12-31-2017

# Conservation and Variation of DNA Methylation in Lactuca sativa and Lactuca serriola

Trudi A. Baker
*University of Massachusetts Boston*

# CONSERVATION AND VARIATION OF DNA METHYLATION IN *LACTUCA SATIVA* AND *LACTUCA SERRIOLA*

A Dissertation Presented

by

TRUDI A. BAKER

Submitted to the Office of Graduate Studies,
University of Massachusetts Boston,
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2017

Biology Program

CONSERVATION AND VARIATION OF DNA METHYLATION IN

*LACTUCA SATIVA* AND *LACTUCA SERRIOLA*

A Dissertation Presented

by

TRUDI A. BAKER

Approved as to style and content by:

_____
Richard V. Kesseli, Professor
Chairperson of Committee

_____
Linda Huang, Professor
Member

_____
Kellee Siegfried, Assistant Professor
Member

_____
David Weisman, Vice President
Indigo Ag, Inc.
Member

_____
Gregory Beck, Program Director
Biology Program

_____
Richard V. Kesseli, Chairperson
Biology Department

# ABSTRACT

## CONSERVATION AND VARIATION OF DNA METHYLATION IN *LACTUCA SATIVA* AND *LACTUCA SERRIOLA*

December 2017

Trudi A. Baker, B.S., Cornell University
J.D., Suffolk University
Ph.D., University of Massachusetts Boston

Directed by Professor Richard V. Kesseli

Molecular techniques for guiding plant breeding have successfully used wild progenitors of domestic crops as sources of genetic variants conveying desirable traits. However, epigenetic variation, in particular DNA methylation, is a significant source of phenotypic variation and epigenetic effects of plant domestication are poorly understood. Described herein are the first single-base pair resolution methylomes of the highly valued crop iceberg lettuce (*Lactuca sativa* cv. Salinas) and its close relative, and ubiquitous weed, *L. serriola*. This work suggests several roles for acquisition and inheritance of methylation in the evolution of *Lactuca* spp. in response to stress. The *Lactuca* spp. have conserved patterns of methylation around genomic regions associated with biotic stress response and conserved changes in average methylation levels in genic and intergenic regions under nutrient deprived

conditions. The genotypes also have important differences in both methylation levels and variability in both control and nutrient deprived conditions. Additionally, there are suggestions that abiotic stress associated methylation may be transmitted between generations with fidelity. Together these findings suggest an additional source and mechanism of genomic variation which may be isolated and adapted for improvement of crops.

ACKNOWLEDGEMENTS

Sincere thanks to Richard V. Kesseli for an extraordinary opportunity; to Bonnie A. Gulick, Jonathan D. Baker, Susan M. Huse for endless faith and encouragement; to Stuart H. Morey, Amy Avery, Michie Yasuda for generously sharing your expertise; and to David Weisman for introducing me to Python, which substantially improved the course of my career.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

| | |
|---|---|
| Amplified fragment length polymorphism | AFLP |
| Argonaute 4 | AGO4 |
| Adenosine tri-phosphate | ATP |
| Base pair | bp |
| Chromomethylase 3 | CMT3 |
| Demeter | DME |
| Demeter-like protein 2 | DML2 |
| Demeter-like protein 3 | DML3 |
| Deoxyribonucleic acid | DNA |
| Dicer-like 3 | DCL3 |
| Differentially methylated cytosine | DMC |
| Differentially variable cytosine | DVC |
| Domains Rearranged DNA Methyltransferase 2 | DRM2 |
| False discovery rate | FDR |
| Flagellin-sensitive 2 | FLS2 |
| First self-ed generation | S1 |
| Homologous recombination frequency | HRF |
| Individuals grown for two generations without fertilizer | NN |
| Individuals grown in nutrient deprived conditions whose parents were grown in control conditions | CN |
| Individuals grown for two generations in control conditions | CC |
| Individuals grown in control conditions whose parents were grown in nutrient deprived conditions | NC |
| Leucine-rich repeat | LRR |
| Long terminal repeat | LTR |
| Nucleotide | nt |
| Messenger ribonucleic acid | mRNA |
| Methylation sensitive amplified polymorphism | MSAP |
| Methylcytosine | mC |
| Methyltransferase 1 | MET1 |
| Mutation accumulator | MA |
| Mutator-like transposable element related locus | MULE-F19G14 |
| Reduced representation bisulfite sequencing | RRBS |
| Repressor of silencing 1 | ROS1 |
| Restriction fragment length polymorphism | RFLP |
| Ribonucleic acid | RNA |
| RNA-dependent RNA polymerase 2 | RDR2 |
| RNA-directed DNA methylation | RdDM |
| RNA polymerase V | Pol V |
| RNA polymerase IV | Pol IV |
| Sawadee Homeodomain Homologue 1 | SHH1 |
| Single nucleotide polymorphism | SNP |

| | |
|---|---|
| Suppressor of variegation 3-9 homolog protein 2 | SUVH2 |
| Suppressor of variegation 3-9 homolog protein 9 | SUVH9 |
| Transcription start site | TSS |
| Transcription end site | TES |
| Variable In Methylation | VIM |
| Whole genome bisulfite sequencing | WGBS |

# CHAPTER 1

# INTRODUCTION

## Epigenetic diversity and its importance

To convey desirable traits such as increased yield or disease-resistance into domestic crops, plant breeders have traditionally used wild relatives of crops as sources of genetic variants conveying desirable traits [1–3]. This has been particularly effective in the many high value agricultural crops closely related to hardy weeds [2,3]. Improvement methods are traditionally targeted toward identification of discrete alleles to improve crops by selective breeding or transgenics [4,5]. However, recent studies have shown significant phenotypic variation associated with epigenetic diversity [6–9].

Epigenetics is broadly defined as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" [10]. Epigenetic modifications can include chemical modifications of nucleotides including methylation of cytosine producing 5-methylcytosine, the "fifth base" [11]. Epigenetic modifications also include chemical modification of proteins closely associated with genomic DNA, most notably the post-translational modification of histone proteins, as well as mitotically heritable protein-DNA associations [12]. Even conformation changes of prion proteins and cytoplasmic inheritance, transmission of

plastids, endosymbionts, viruses, and small RNAs through mitosis or meiosis, are sometimes broadly considered epigenetic [13–16]. Indeed, many of these concepts are interrelated. For example, DNA methylation in the CHG (H is any nucleotide but G) context positively reinforces the histone modification Histone3 Lysine9 [17]. The genome of the symbiotic plant root endophyte *Mesorhizobium loti* undergoes adenine methylation in the process of symbiosis, and these modifications are required for the efficient formation of nodules on the plants' roots [18]. Even the stalwart of traditional agricultural breeding programs, heterosis or hybrid vigor, has an epigenetic component [19,20].

**Key epigenetic marks and mechanisms**

DNA methylation will be the focus of this chapter and the following chapters. Though DNA methylation is one of many epigenetic mechanisms, it is highly prevalent, central to the interaction of other epigenetic mechanisms, and has unique characteristics in plant genomes. In plants, considerable methylation is found in each of the three possible sequence contexts for methylcytosine: CG, CHG and CHH, where H is any nucleotide except G. In plants, as in most eukaryotes, methylation is most frequently found in the CG context [21]. In differentiated human fetal fibroblasts more than 99.98% of methylation is found in the CG context [22], whereas in Arabidopsis immature floral tissue only a slight majority (55%) of methylcytosines are found in the CG context [23]. An additional differentiating characteristic of mammalian and plant epigenomes is the degree to which acquired DNA methylation is reset between generations. Methylation in

mammalian genomes goes through two round of erasure during embryogenesis, whereas methylation in plant genomes, particularly in the CG and CHG contexts, are maintained through embryogenesis [24].

Methylation in different plant genomes share some common characteristics. Methylation within the CG context is more frequent than CHG or CHH within gene bodies [23,25–30]. In *Arabidopsis thaliana*, when the MET1-3 methyltransferases were knocked out, CHG methylation in euchromatic regions increased significantly and CHG methylation within gene bodies was enriched, taking on a similar profile to that of CG in wild-type plants [23]. Given the interrelation of the methylation and siRNA pathways, it is not surprising that Lister et al. (2008) found 85% of genomic regions with small RNA sequence identity contained at least one methylated cytosine; those methylation sites comprised 39% of all methylated sites [23].

**Methylation polymorphisms and plant phenotypes**

Methylation polymorphisms can be introduced stochastically due to a lack of fidelity of DNA methylation maintenance or by presence of a genetic variant such as a repetitive element insertion, or as a targeted response to environmental stimuli. Schmitz et al. (2011) estimated the rate of methylation polymorphism per CG to be 100,000 times greater than the rate of per nucleotide sequence polymorphism for the mutation accumulator (MA) lines of *A. thaliana.* This rate was based upon 30 generations derived from common ancestry. However, sites of methylation polymorphism between MA lines were not evenly distributed through the genome rather they were concentrated in certain

genomic locations [31,32]. This concentration of methylation polymorphism was also seen between members of geographically disparate wild populations of *A. thaliana* having diverged over a century ago [33]. This suggests that the fidelity of DNA methylation maintenance in plants may be highly variable across the genome, as has been shown in mouse embryonic stem cells [34].

Methylation affects plant phenotypes important for reproduction and fitness, including flower morphology [35], flowering time [36–38], sex determination [39], herbivore and pathogen resistance, [40–42] and agronomically important traits such as heterosis [19]. In general, abiotic stress associated methylation results have not been consistently mitotically or meiotically transmitted and, when detected, the direction of methylation change has been, in some cases, inconsistent across species and conditions. Treatment of *Zea mays* seedlings with intermittent heat, cold and ultraviolet stresses, for instance, did not result in condition-specific methylation patterns in adult plants [43], nor did methylation changes in rice correlate with salt treatment or the salt tolerance of the variety [44,45]. A recent study found both salt sensitive and resistant varieties of rice were globally hypomethylated in salt treatments [45]. In contrast, salt stressed *A. thaliana* were globally hypermethylated [46]. The duration of the stress and the length of time between stress treatments and tissue sampling may be important, but largely unconsidered, variables when comparing different stress methylation studies; methylation changes can be induced within a few hours of stress treatment [45,47] and a large proportion of induced changes may revert with time [48].

Methylation can affect phenotype through many different mechanisms including modulating gene expression, transposable element mobility, and alternative splicing patterns. The relationship between methylation levels and gene expression levels differs by gene region. Methylation levels within the promoter regions are traditionally thought of as negatively correlated with gene expression levels through decreased binding affinity of transcription factors for methylated DNA [49] and reduced access of transcription factors and binding sites due to methylation-induced compact chromatin structure [50]. An example of negative correlation between promoter methylation and gene expression, is the expression of key genes involved in ethylene-induced ripening in tomatoes which requires active demethylation of their promoter regions by DNA glycosylases SlDML1/2 (orthologs of ROS1 in Arabidopsis) [27,51]. Interestingly, ROS1 itself is an important counter example to the generally inverse relationship between promoter methylation and gene expression. The ROS1 promoter is the target of both RdDM and active demethylation by ROS1 and transcription of the ROS1 gene is directly and positively correlated with the methylation levels in its promoter [52]. Thus the ROS1 promoter acts as a self-regulating rheostat, maintaining balance in its own methylation levels as well as ROS1 targets such as transposable element proximal genes [52]. Contrary to promoter regions, moderate methylation levels within gene bodies are associated with highly expressed genes, while high and low levels of methylation are associated with low expression levels [53]. Though methylation levels in both promoter and gene bodies are associated with expression, only a small percentage of differentially expressed genes in Arabidopsis and rice are associated with differential methylation [54], and only ~20% of

maize genes with on-off expression differences between inbred maize lines were associated with differentially ethylated regions [55].

Gene body methylation has also been associated with the suppression of transposable element insertion [56] and alternative splicing [57]. Mutator transposons insert preferentially into unmethylated regions [56,58], preferentially into genes, and more specifically into regions depleted in CG, but not CHG or CHH methylation [58]. The high average levels of CG methylation found in the gene bodies of most angiosperms could be an adaptive defense to transposable element insertion. Gene body methylation is also associated with prevalence of isoforms. Regulski (2013) analyzed sites of alternative splicing and found a bias in acceptor sites toward lower levels of CHG methylation, while levels of CHH methylation did not appreciably affect splicing efficiency. For honey-bee genes that are alternatively spliced, skipped exons are significantly hypomethylated relative to included exons, though in both cases exons have higher levels of methylation than flanking introns [57].

Transposable elements generally have higher average levels of methylation than intergenic regions [25,59], and methylation can serve to suppress their transcription and mobility [21]. During gametogenesis, passive demethylation and active DNA demethylation by DEMETER are associated with active transcription of transposons in the vegetative and central cells [21]. These transcripts travel to their respective egg or sperm cells and reinforce transcriptional silencing and RdDM of transposons [60]. In somatic tissue, hypomethylation in loss of function methyltransferase mutants in Arabidopsis resulted in a significant increase in transposon and pseudogene transcription

relative to wild-type [23]. A variety of biotic and abiotic stresses are associated with hypomethylation [45,61,62] and treatment of Arabidopsis with the plant stress associated phytohormone salicylic acid resulted in hypomethylation of transposable elements and increased transcription of those elements [40]. However, not all releases of transposable elements are associated with removal of methylation. For example the release of silencing of Mutator-like transposable element related locus (MULE-F19G14) with temperature shifts occurs despite the maintenance of high levels of methylation and repressive histone modifications (H3K9/K27) [63]. And in an additional example, the binding affinity of the Tam3 transposase to its binding site in the sub-terminal repeat region of Tam3 is impaired by DNA methylation *in vitro*, but lack of methylation *in vivo* is not sufficient to induce transposition [64].

The presence and methylation status of transposable elements can affect the expression of proximal genes and potentially alter the organisms' fitness [40,65–67]. For example, genes which are up are upregulated in stress conditions in maize are enriched near certain families of transposable elements [68]. In Arabidopsis, genes which are downregulated in loss-of-function demethylase mutants with increased susceptibility to *Fusarium oxysporum* infection are enriched with transposable element regions in their promoters [69]. Methylation at neighboring transposable elements may have a positive or negative correlation to expression levels of nearby transposable genes [40,70].

**Transgenerational methylation effects**

Specific environmental stresses to a parent can result in identifiable differences in their offspring's DNA methylation levels, their expression of stress-related genes, and their competitive ability in the stress environment [46,71]. In considering the evolutionary consequences of inheritance of acquired methylation, two possible scenarios can be considered. Stress associated DNA methylation could be directed towards the stress that is encountered, as has been suggested in mammalian nutritional studies. Several studies have found that offspring of parents with altered nutritional states (starvation, high fat, low-protein diets) had altered DNA methylation of metabolism related genes [71–73]. And in plants, members of the RdDM pathway play a role in transgenerational priming – the phenomenon where exposure to stress can make the individual or its offspring better poised to respond to future incidents of the stress [42]. Alternately, inheritance of stress associated DNA methylation could be beneficial, not by changing the mean level of methylation at a target, but rather by introducing stochasticity. Modeling supports the hypothesis that stochastic variation in mC would be advantageous in a disturbed environment [74]. Additionally, experimental data in dandelion shows increased epigenetic variation within individuals in stress treatments relative to unstressed groups [8]. In Arabidopsis, groups of genetically identical, but epigenetically diverse individuals, were more resistant to pathogen challenge and competition than less epigenetically diverse groups [75].

**Epigenetic effects of domestication of *Lactuca* species**

Long histories of artificial selection have resulted in significant phenotypic divergence of domestic plants from wild relatives and decreased genetic diversity among selected domestic varieties. The epigenetic consequences of such prolonged selection are unknown. This dissertation seeks to provide a deeper understanding of the directional vs. stochastic hypotheses through a comparative analysis of the methylomes of domestic lettuce *Lactuca sativa* and its closely related wild and weedy relative *L. serriola. L. sativa* and *L. serriola* are particularly well suited for this study as they are self-fertile, populations tend to be highly homozygous. In the absence of genetic diversity, epigenetic diversity may be even more important. Additionally, *L. sativa* and *L. serriola* have very different tolerances for stress. *L. sativa* is commercially produced in a narrow range of environmental conditions and production is nutrient intensive, using more nitrogen fertilizer per acre than corn and most other vegetables [76]. *L. sativa* is also susceptible to many pathogens to which its wild relative *L. serriola* is substantially resistant [77].

The domestication history is well documented. Domestication of *L. sativa* has been traced to the Middle Eastern region encompassing modern day Iraq, Turkey, Syria, Lebanon, Israel, and the Egyptian river valley, referred to as the Fertile Crescent [78,79]. Significant cultivation of *L. sativa* is documented through the Grecian (450 B.C.) and Roman (1 A.D.) empires [78]. The first report of a heading type lettuce, *L. sativa* var. *capitate*, dates to a 1543 herbal book of German horticulturalist Leonhart Fuchs [78]. Though Linneaus (1757) classified *L. sativa* and *L. serriola* as distinct species, the accuracy of this taxonomic distinction and the exact relationship of *L. sativa* to *L.*

*serriola* and other related species has been contentious. As early as 1851, Bischoff, Boissier, Hooker and Fiori contended that *L. sativa* and *L. serriola* were conspecific, differing only in degree of domestication. The relationships of these taxa were resolved with the application of molecular markers. Kesseli et al. 1991 showed that *L. serriola* alone was progenitor of *L. sativa* and that none of the 143 RFLP loci examined had diagnostic alleles that separated these taxa [80]. This was further confirmed with AFLP data [81]. Interestingly however, the major morphological groups of lettuce each appear to have a monophyletic origin suggesting that they arose from independent lineages of *L. serriola*, an idea first proposed by Sturtevant in 1886 [82]. The first report of *L. serriola* in the United States was in L.H. Pammel's 1863 report on distribution of weeds [78]. *L. serriola* is a particularly ubiquitous weed, found on all continents except for Antarctica, and as a common weed found throughout the lower 48 states [83]. *L. serriola* is a hardy weed commonly found beside highways and in other human disturbed environments.

The experiments described in the following chapters suggest several roles for acquisition and inheritance of methylation in the evolution of *Lactuca* spp. in response to stress. The *Lactuca* spp. have conserved patterns of methylation around genomic regions associated with biotic stress response and conserved, stress-induced changes in average methylation levels in genic and intergenic regions. These experiments also offer insights into the role of variability in methylation in the genotypes' differing response to stress conditions. Additionally, there are suggestions that abiotic stress associated methylation may be transmitted between generations with fidelity. Together these findings suggest an additional source and mechanism of genomic variation which may be isolated and

adapted for improvement of crops. Chapter 2 describes the first whole genome bisulfite sequencing of *Lactuca* species, and examines the differences in methylation patterns between *L. sativa* and *L. serriola* and other plant species through the lens of domestication. Chapter 3 introduces a novel method of characterizing whole genome patterns using reduced representation bisulfite sequencing. Chapter 4 looks at the impact of stress on the acquisition and the transmission of DNA methylation. Finally, Chapter 5 summarizes significant differences between the methylomes of these closely related species and suggests roles for DNA methylation in adaptation to environmental disruption.

# CHAPTER 2

# CONSERVATION AND VARIATION IN DNA METHYLATION IN *LACTUCA SATIVA* AND *L. SERRIOLA*

## Introduction

To convey desirable traits such as increased yield or disease-resistance into domestic crops, plant breeders have traditionally used wild relatives of crops as sources of genetic variants conveying desirable traits [1–3]. This has been particularly effective in the many high value agricultural crops closely related to hardy weeds [2,3]. Improvement methods are traditionally targeted toward identification of discrete alleles to improve crops by selective breeding or transgenics [4,5]. However, recent studies have shown significant phenotypic variation associated with epigenetic diversity [6–9].

Methylation affects plant phenotypes important for reproduction and fitness, including flower morphology [35], flowering time [36–38], sex determination [39], herbivore and pathogen resistance, [40–42] and agronomically important traits such as heterosis [19]. In general, abiotic stress associated methylation results have not been consistently mitotically or meiotically transmitted and, when detected, the direction of methylation change has been, in some cases, inconsistent across species and conditions.

Treatment of *Zea mays* seedlings with intermittent heat, cold and ultraviolet stresses, for instance, did not result in condition-specific methylation patterns in adult plants [43], nor did methylation changes in rice correlate with salt treatment or the salt tolerance of the variety [44,45]. A recent study found both salt sensitive and resistant varieties of rice were globally hypomethylated in salt treatments [45]. In contrast, salt stressed *A. thaliana* were globally hypermethylated [46]. The duration of the stress and the length of time between stress treatments and tissue sampling may be important, but largely unconsidered, variables when comparing different stress methylation studies; methylation changes can be induced within a few hours of stress treatment [45,47] and a large proportion of induced changes may revert with time [48].

Specific environmental stresses to a parent can result in identifiable differences in their offspring's DNA methylation levels, their expression of stress-related genes, and their competitive ability in the stress environment [46,71]. In considering the evolutionary consequences of inheritance of acquired methylation, two possible scenarios can be considered. Stress associated DNA methylation could be directed towards the stress that is encountered, as has been suggested in mammalian nutritional studies. Several studies have found that offspring of parents with altered nutritional states (starvation, high fat, low-protein diets) had altered DNA methylation of metabolism related genes [71–73]. And in plants, members of the RNA directed DNA Methylation (RdDM) pathway play a role in transgenerational priming – the phenomenon where exposure to stress can make the individual or its offspring better poised to respond to future incidents of the stress [42]. Alternately, inheritance of stress associated DNA

methylation could be beneficial, not by changing the mean level of methylation at a target, but rather by introducing stochasticity. Modeling supports the hypothesis that stochastic variation in mC would be advantageous in a disturbed environment [74]. Additionally, experimental data in dandelion shows increased epigenetic variation within individuals in stress treatments relative to unstressed groups [8]. In Arabidopsis, groups of genetically identical, but epigenetically diverse individuals, were more resistant to pathogen challenge and competition than less epigenetically diverse groups [75].

Many biotic stresses are associated with global or loci specific hypomethylation. Global loss of methylation is associated with increased resistance to infection by the bacterial pathogen *Pseudomonas syringae* in Arabidopsis [40] and upregulation of stress response genes in transgenic tobacco (*Nicotiana tabacum* cv. Xanthi) [47]. Treatment of rice with the methyltransferase inhibitor 5-azadeoxycytidine induced global hypomethylation and resistance to infection by *Xanthomonas* [84]. In addition, DNA glycosylase loss of function mutants have been shown to be more susceptible to fungal and bacterial pathogens and showed increased methylation and decreased expression of stress response genes [41,69]. Plants have evolved proteins, encoded by resistance genes, which recognize effector proteins of pathogens resulting in elicitor-triggered immunity [85,86].

The most prevalent class of resistance genes in plants are NBS-LRR proteins, which contain a leucine-rich repeat (LRR) domain [87]. Boyko et al. (2007) found significant hypomethylation and increases in homologous recombination frequency (HRF) in LRR domain containing genes in tobacco plants challenged with tobacco

mosaic virus [62]. A key protein in the recognition of pathogen infection is the plant pattern-recognition receptor FLAGELLIN-SENSITIVE 2, it recognizes bacterial flagellin-derived peptide 22 (flg22). Flg22 triggers active demethylation by DNA glycosylase ROS1 and upregulation of some long terminal repeat (LTR) containing transposable elements and some LRR containing resistance genes [41]. In the absence of pathogen pressure, ROS1 constitutively demethylates these transposable elements in balance with constitutive transcriptional gene silencing. Pathogen pressure in a loss of function ROS1 mutant resulted in aberrant methylation in the CHH context of ROS1-target LTR transposable elements [41].

Fungal and bacterial pathogens cause significant losses in the production of lettuce (*L. sativa*), the most consumed vegetable in the United States whose annual production is valued at approximately $2 billion [76]. A close relative, and ubiquitous weed, *L. serriola* shows enhanced resistance to many of these pathogens. Early molecular work showed that *L. serriola* is the sole progenitor of *L. sativa*, supporting the contention that fully cross fertile *L. sativa* and *L. serriola* are conspecific [80,81]. Breeding efforts to enhance pathogen resistance in *L. sativa* include introduction of resistance genes from *L. serriola* [77,88–92]. In this paper we explore differences in the methylomes of *L. sativa* cv. *Salinas*, one of the mostly widely used elite cultivars in the breeding of modern crisphead lettuce varieties [93], and *L. serriola* (UC96US23), a pervasive and hardy weed, in the context of domestication and pathogen resistance phenotypes.

## Methods

*Samples and Extraction*

*Lactuca sativa* cv. Salinas and *L. serriola* (UC96US23) seeds were obtained from Richard Michelmore's lab at the University of California Davis and the Compositae Genome Project (http://compgenomics.ucdavis.edu/). To reduce variation due to the maternal effect of different growing conditions for the different sources of seeds used in this study, two "progenitor" generations were planted (procedure described below). To avoid individual specific maternal effects, offspring from different self-fertilized parent plants were used as biological replicates. Seeds of each genotype were sterilized according to the following procedure: 1 mL 20% bleach solution and one drop Tween 20 were added to 25 seeds in a 2 mL microcentrifuge tube and gently agitated for 5 minutes. After a quick spin, detergent solution was decanted and 1 mL autoclaved, deionized water added and tubes gently agitated for 5 minutes.

This process was repeated for a total of 10 rinses. The seeds were refrigerated overnight at 4°C. Seeds were planted in commercial potting soil (Fafard Growing Mix 2: 70% Canadian sphagnum peat, 30% perlite and vermiculite) that had been autoclaved (25 minutes wet cycle) in two consecutive days preceding planting. The autoclaved soil was thoroughly moistened with autoclaved deionized water prior to filling half-gallon nursery pots. Sterilized seeds were then planted 2 seeds per container, at approximately 6 mm depth, covered with aluminum foil, then refrigerated at 4°C for 5 days.

Plants were randomly assigned positions within a 72 square grid in a Coviron® PGW36 Plant Growth Chamber at the University of Massachusetts Boston. Standard growth conditions were 16 hours of 800 µmol/$m^2$/s intensity light at 23°C and 8 hours dark at 18°C. For the first two weeks in the growth chamber plants were watered 6 days per week with autoclaved deionized water. Thereafter plants were watered 2 times per week with unamended autoclaved deionized water and once with autoclaved deionized water supplemented with Peter's 20-20-20 all-purpose fertilizer at a concentration of 120 parts per million (N).

Tissue was collected from four biological replicates of *L. sativa* and *L. serriola*. In order to minimize variation between samples due to developmental differences, leaf tissue was collected when the first individual flowers of the secondary inflorescence are visible but still closed [28]. Samples were collected at a consistent time of day, between 1 and 2 hours prior to daybreak, to minimize variation in stress-related transcriptomes [94,95]. Two, 13 mm diameter leaf discs were placed in sterile containers and immediately immersed in liquid nitrogen.

*Whole genome bisulfite library preparation and sequencing*

DNA extractions were performed using MoBio's PowerPlant Pro DNA extraction kit with the following modifications: 40 µl of Phenolic Separation Solution was added to 410 µl of Solution PD1, 50 µl Solution PD2 and 3 µl RNaseA samples were added, then incubated at 65°C for 10 mins. Samples were further purified using MoBio's PowerClean Pro DNA Clean-up kit accord to manufacturer's instructions.

For each sample 1.4 µg DNA was fragmented using the Covaris S220 system (Covaris, Woburn, MA) in microTUBE AFA Fiber tubes (Covaris, Cat. No. 520045) to between 100 and 500 bp using the following instrument parameters: 80 s with a duty factor of 10%, a peak incident power of 175W, a temperature of 6°C and 200 cycles per burst. The sonicated DNA was purified using Qiagen DNeasy MinElute columns according to manufacturer's instructions. End-repair and ligation were performed using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® and NEBNext® Multiplex Oligos for Illumina® (Methylated Adaptors) according to manufacturer's instructions (New England Biolabs, Ipswich, MA). Ligation products were purified using a 1:1 ratio of Agencourt AMPure XP beads (Beckman Coulter Genomics, Danvers, MA) to product, and eluted in 30 µl of 1 x TE buffer. Libraries were size selected with a 1.5% Blue Pippin agarose gel cassette (Sage Science, Beverly, MA) for fragment sizes between 250 and 600 bp to collect fragments with a minimum 180 bp insert size plus the additional length of the ligated adapters. Isolated products were purified using a 1:1 ratio of Agencourt AMPure XP beads to product and eluted in 10 µl nuclease-free water. Bisulfite conversion was performed using NEB's Epimark Bisulfite Conversion kit according to manufacturer's instructions with an additional 5:1 bead clean-up.

Each WGBS library was PCR amplified in three separate 50 µl reactions, each containing 6.65 µl of bisulfite sample, 1 µl 10 mM dNTPs, 0.75 µl NEBNext universal PCR primer, 0.75 µl NEBNext index primer, 10 µl 5X EpiMark hot start Taq reaction buffer, 0.25 µl NEB EpiMark hot start Taq DNA polymerase and 30.6 µl ultrapure water. PCR conditions were 95°C for 30 seconds, followed by 13 cycles of 95°C for 15 seconds,

61°C for 30 seconds, and 68°C for 30 seconds, a final extension at 68°C for 5 minutes and hold at 4°C. Following PCR, the triplicate samples were pooled prior to bead purification. PCR products were purified using a 0.79:1 ratio of Agencourt AMPure XP beads to product, immediately followed by a subsequent bead purification using a 1:1 ratio and eluted in 20 µl 10 mM Tris, pH 8.0. Paired-end sequencing (2x100) was performed on a HiSeq 2000 at the University of Massachusetts Boston, Center for Personalized Cancer Therapy Genomics Core.

Raw reads were converted from bcl to fastq format using bcl2fastq (v. 1.8.4). Reads were trimmed using Trimmomatic (v 0.32) [96]. TruSeq3-PE adapter sequences were used as a reference for adapter trimming, allowing 2 nt mismatches with the adapter sequences, palindrome clip threshold of 30, simple clip threshold of 10. Overlapping paired end reads were merged using leeHom with the alignment of each library to the chloroplast genome serving as a prior [97]. The priors were estimated for each library based on paired end alignment of trimmed (but not merged) reads to the lettuce chloroplast genome.

The rate of bisulfite non-conversion was estimated by aligning paired reads to the bisulfite converted and bowtie2 indexed lettuce chloroplast genome (Accession number: NC_007578.1 by Bismark (v 0.13.1) [98]. Alignments were carried out in a two-step process. First, paired reads were aligned, with the –un option selected. Second, the reads which did not produce a valid, paired alignment were aligned in single read mode. Duplicate reads were marked and removed using Picard Tools' MarkDuplicates (v 1.96) [99]. The resulting bam files were then converted to the sam format using Picard Tools'

SamFormatConverter. Bismark's methylation_extractor script was run with the –bed_graph option which generates a 1-based report ("coverage" file) with the counts of the methylated and unmethylated reads detected at each position and summarizes these results over the entire genome. The coverage files were used as input to Bismark's coverage2cytosine which generates a text file summarizing the counts of methylated and unmethylated reads at each position in the genome regardless of whether any reads covered that position. All samples have apparent chloroplast methylation rates less than 5%.

*Genome preparation, read alignment, and preliminary methylation counts*

Access to the genome assemblies of *L. sativa* (v6) and *L. serriola* (v6) were generously provided by the Compositae Genome Project (http://lgr.genomecenter.ucdavis.edu; S. Reyes-Chin Wo, A. Kozik, D. Lavelle, and R.W. Michelmore, unpublished data). Sequences were bisulfite converted using Bismark's bismark_genome_preparation.

Trimmed reads were aligned to the bisulfite converted and indexed genome using Bismark [98] and bowtie2 [100]. Alignments were generated in the two step process described above, except the leeHom merged reads were also aligned as single reads in the second step. The resulting genome alignment files for each biological replicate were combined by replicate, and duplicate reads were marked and removed using Picard Tools' MarkDuplicates [99].

Bismark's methylation_extractor script was run with the –bed_graph option which generates a 1-based report ("coverage" file) with the counts of the methylated and

unmethylated reads detected at each position and summarizes these results over the entire genome. The coverage files were used as input to Bismark's coverage2cytosine which generates a text file summarizing the counts of methylated and unmethylated reads at each position in the genome regardless of whether any reads covered that position.

*Comparison of methylation between L. sativa and L. serriola by genomic feature*

To calculate summary statistics for percent methylation by context over coding and repetitive regions, we summed the number of methylated reads and total reads of the biological replicates aligned to their respective genome sequences, selected only positions which were covered by five reads in all replicates of a genotype, and calculated the percent methylation by combining all replicates. These results were saved in the bed format and bedtools intersect [101] was used to define positions based on feature and feature proximity. Gene features were limited to those predicted loci also having transcriptional support and filtered to only the primary transcript per locus to avoid double counting. Repeat features were identified using RepeatMasker v4 [102].

Average methylation levels over protein coding genes and surrounding up- and downstream regions were calculated based on approximately 37,000 predicted genes in these genomes. Average levels were calculated over 100 bp bins in regions 10,000 bp upstream of the transcription start site (TSS), and 10,000 bp downstream of the transcription end site (TES) for all protein coding genes. Protein coding genes greater than 1000 nt in length were divided into 100 bins, and summary statistics computed for each bin position.

To detect differential methylation between *L. sativa* and *L. serriola*, reads from *L. serriola* were aligned to the *L. sativa* (v6) genome as described above. Only genome positions with at least three reads in each of the four *L. sativa* and *L. serriola* replicates and positions with non-zero variance in the proportion of methylated reads were retained for further analysis. Reads were analyzed in R using MethylSig [103]. Local information was included in the estimation of variance but not local methylation level. The local dispersion level was calculated across 8-9 orders of magnitude and repeated for tiled regions of: 1 bp, 10 bp, 100 bp, 1000 bp for each context. The smallest dispersion window that maximized detection was 1 Mb, corresponding to a window of +/- 1.17 cM. The differentially methylated cytosines (DMCs) with q-value <0.05 and a methylation difference >= 20% were considered significant. The predicted protein coding genes and repetitive features which overlapped with these DMC were identified using bedtools intersect.

*Detection of differentially variable methylation between L. sativa and L. serriola*

We utilized the iEVORA algorithm to test the null hypothesis of equal variances between biological replicates of *L. sativa* and *L. serriola* in the proportion of methylation at cytosines covered by at least ten reads using a q-value threshold of 0.001 [104,105]. The predicted protein coding genes and repetitive features which overlapped with these DMC were identified using bedtools intersect.

Gene ontology and KEGG annotations obtained from the Compositae Genome Project were used to perform gene ontology analysis of genes in differentially methylated and differentially variable regions. Hypergeometric testing using the R package phyper

with false discovery rate (FDR) correction ($p < 0.05$) was performed to detect gene ontology terms over-represented in the entire set of genes containing one or more DMC, and for the sets located within 1 kb or 2 kb of an annotated repetitive element.

## Results

### *Whole genome bisulfite sequencing of L. sativa and L. serriola*

We performed whole genome bisulfite sequencing of four *L. sativa* and four *L. serriola* individuals, obtaining an average of 52 million high quality, deduplicated reads per individual. In *L. sativa*, the average methylation percent was 84.2% in the CG context, 70.3% in the CHG context, and 12.6% in the CHH context (Figure 1A). In *L. serriola*, average methylation levels were slightly lower in all sequence contexts; 77.4%, 64.5% and 10% respectively (Figure 1B). These methylation levels were comparable to other plant genomes of similar size (Figure 2A).

Among plant species the proportion of methylation at cytosines in the CG and CHG positions are significantly and positively correlated with genome size (CG: $R^2$=0.57, p-value=0.0046 and CHG: $R^2$=0.78, p-value=0.0002), whereas the relationship between methylation levels in the CHH context and genome size is weak and not statistically significant ($R^2$=0.15 and p-value=0.2149; Figure 3).

Both *L. sativa* and *L. serriola* have the characteristic bi-modal distribution of methylation levels with the vast majority (99.03% CG, 91.77% CHG, 90.14% CHH) of cytosine positions less than 20% or more than 60% methylated. In all sequence contexts the coefficient of variation for percent methylation at a position is inversely related to the

average percent methylation (Figure 4). These results are consistent with the observation in wild and domestic rice [54] suggesting an evolutionary role in conservation of highly methylated positions particularly those inhibiting the spread of transposable elements [60].

Previous work in other plants species has shown significant correlation of methylation levels for up to 5 kb [25]. We found the spatial autocorrelation of methylation in all contexts was highly consistent between all *L. sativa* and *L. serriola* replicates (Figs 5 and 6), though the degree of correlation differed significantly between the CG/CHG and the CHH context. The average correlation of methylation levels between cytosines separated by up to 50 kb is 0.82 in the CG context, 0.70 in the CHG context and 0.22 in the CHH context. The correlation between positions in CG or CHG contexts decreases after 50 kb, whereas the correlation does not vary significantly with genomic distance in the CHH context.

Methylation levels in both *Lactuca* ssp. took on familiar patterns in the regions up- and downstream of protein coding genes patterns where the relatively high levels of methylation in the CG and CHG contexts decrease sharply in the regions preceding the transcription start site, increase over the gene body, decrease towards the 3' end of the transcribed region, then increase in the downstream region. These patterns are very similar to those previously shown in *A. thaliana* [23,25] and other *Brassicaceae* [30], *Oryza. sativa* [26,54], *Populus trichocarpa* [26], *Manihot esculenta* [106], *Glycine max* [107], *Solanum lycopersicum* [27], and *Zea maize* [28] (Figure 6). *Lactuca* also shows increases in CHH methylation within a few hundred nucleotides upstream of the

transcription start site and downstream of the poly-adenylation signal similar to previous reports in *Zea maize* [28] (Figure 7).

The average methylation percentage over 1,063 of the resistance genes in *L. sativa* [108] showed strikingly different patterns of methylation compared to the trends observed when considering all protein coding genes. In both *L. sativa* and *L. serriola*, the regions from 100 to 400 bp upstream and downstream of resistance genes were substantially more methylated in the CHH context compared to the average for other genes (Figure 8 A and B). The methylation percentage over the resistance genes themselves were low in all contexts with significant spikes at the 3' end of the genes (Figure 8 C).

*Identifying differentially methylated cytosines between L. sativa and L. serriola*

To identify cytosines with significant differences in mean methylation level between *L. sativa* and *L. serriola*, we aligned reads from biological replicates of *L. sativa* and *L. serriola* to the *L. sativa* genome sequence. We filtered the aligned positions and considered only positions covered by at least three reads in all replicates of both genotypes and having non-zero variance in the proportion of methylated reads between genotypes. There were 293,264 such cytosines in the CG, 257,984 in CHG, and 1,392,713 in CHH contexts. We tested for differential methylation using a beta binomial model across biological replicates [103]. Cytosine sites for which *L. sativa* and *L. serriola* had significant differences in methylation levels ($q < 0.05$) and also differed by at least 20% in mean methylation levels were considered for further analysis. There were

5,344 differentially methylated cytosines (DMCs) in the CG, 3,909 in the CHG, and 3,306 in the CHH contexts. Of these positions, 1,064 (7.81%) were associated with known sequence polymorphism between *L. sativa* and *L. serriola* and were excluded from further analyses of DMCs.

The mean level of methylation at DMCs was higher for *L. sativa* in all three contexts; 70% vs. 42% in the CG context, 67% vs. 34% in the CHG context, and 50% vs. 36% in the CHH context. In pairwise comparisons of these DMCs, *L. sativa* had the higher methylation level in 72%, 76% and 62% of positions in the CG, CHG and CHH contexts, respectively. These cytosines mostly were found in repetitive or unannotated regions (Figure 9 A-C). Most DMC within annotated repetitive elements are found in LTR retrotransposons (Figure 9 D-F).

We located 740 DMCs in 318 genes; 357, 282, and 101 DMCs in the CG, CHG and CHH contexts respectively. The majority (67.84%) of DMCs within genes had higher methylation levels in *L. sativa* than *L. serriola*. Genes with the highest frequency of DMC in their gene bodies were annotated with terms including hydroxylase activity, monooxygenase activity, electron carrier activity, transmembrane receptor activity, metal ion transport, and protein kinase activity (Table 1). The gene ontology terms including transmembrane receptor activity, intrinsic to membrane, serine-type endopeptidase activity, quinone binding, and oxidoreductase activity were enriched in DMC containing genes relative to their occurrence in the entire set of genes in the genome (Table 2).

As methylation in upstream regions of resistance genes in close proximity to repetitive elements has been associated with pathogen resistance [40,69], we looked at

DMCs in genes and regions 1 kb regions upstream that were also within 1 kb of a predicted repetitive element. Approximately half of all DMCs in genes, and 18% of DMCs in regions 1 kb upstream of genes, are located within 1 kb a predicted repetitive element. Of DMCs in genes located within 1 kb of a predicted repetitive element, 67% were more highly methylated in *L. sativa*; 83% of DMCs in upstream regions within 1 kb of a predicted repetitive element had a higher percent methylation in *L. sativa*. Methylation levels at DMC were significantly higher in *L. sativa* ($p < 0.05$) in the CG and CHG contexts, but did not significantly differ in the CHH context. Though the average methylation levels at DMC differed between *L. sativa* and *L. serriola*, within *L. sativa* or within *L. serriola* the level of methylation in upstream or gene regions was not significantly different between features located within 1 kb of a repetitive element and those greater than 1 kb from a repetitive element. Though methylation levels at DMC within upstream regions were significantly higher in *L. sativa* than *L. serriola,* methylation levels were not affected by proximity to repetitive regions (Figure 11 C and D). Gene ontology terms enriched in both genes and upstream regions within 1 kb of a predicted repetitive element included serine-type endopeptidase activity, isomerase activity, endonuclease activity, plastid, and carbon fixation, (Table 3 & Table 5).

*Identifying differentially variable methylated cytosines between L. sativa and L. serriola*

We utilized the iEVORA algorithm to test the null hypothesis of equal variances between biological replicates of *L. sativa* and *L. serriola* in the proportion of methylation at cytosines covered by at least ten reads using a q-value (FDR) threshold of 0.001

[104,109]. Very few positions (1.11%) co-localized with known sequence polymorphism and these were excluded from further analysis. In all sequence contexts the variability of methylation was greater in *L. sativa*: 55% of differentially variable methylated cytosines (DVCs) were more variable in *L. sativa* in the CG context, 77% in the CHG context and 92% in the CHH context. There were 378 DVCs in the CG context, 139 in the CHG context and 1,180 in the CHH context. Like DMCs, DVCs were mostly found in repetitive regions and unannotated regions, though only 3.71% of all DVCs were also DMCs and 0.5% of all DMCs also DVCs. Most DVCs within annotated repetitive elements were found in LTR retrotransposons (Figure 12 D-F). DVCs in all three sequence contexts were found within predicted protein coding genes (Figure 12 A-C). Genes containing DVCs were enriched for gene ontology terms structural constituent of ribosome, cytochrome-c oxidase activity, iron ion binding, ribosome, and mitochondrial electron transport cytochrome-c to oxygen (Table 7).

To investigate the spatial correlation of DMCs, DVCs, sequence polymorphisms, annotated genes and annotated repetitive elements, we divided each chromosome into 100,000 equally sized regions and counted the features starting within that region. The regions ranged from 2,271 to 4,369 bp in length depending on the chromosome. The number of DMC within a region was not strongly or significantly correlated with the number of SNPs, protein coding genes, or repetitive elements in that region (Figure 13A). The frequencies of DMCs in the three contexts are weakly, but significantly, correlated with each other (tau=0.18-0.22, p-value«0.01, representative values for chromosome 1). Similarly, the frequencies of DVCs in the three contexts are weakly, but significantly,

correlated (tau=0.48-0.49, p-value « 0.01, representative values for chromosome 1). The frequencies of DVCs in the three contexts are also weakly, but significantly, correlated with DMC in the regions (tau=0.17-0.3, p-value«0.01, representative values for chromosome 1) (Figure 13 B, values includes all chromosomes).

*Gene ontology analysis of conserved methylation*

We defined positions of highly conserved methylation as positions where the variability of percent methylation between biological replicates was in the lowest 25% for that *Lactuca* spp. and the average methylation level of *L. sativa* and *L. serriola* differed by less than 20%. In all sequence contexts, most sites of conserved methylation were located in regions without an annotated protein coding gene or repetitive element (Figure 14). The majority of positions with highly conserved methylation levels had low levels of methylation (Figure 16), and were found in non-repetitive intergenic regions (Figure 15). There were 249,770 cytosines with highly conserved methylation states, 74% of which had average methylation levels of less than 20%, 19.2% had average methylation levels greater than 60%. The majority of positions with conserved low levels of methylation were in the CHH context (80%), with 10.3% and 9.8% in the CG and CHG contexts respectively. In contrast, 51.9% of positions with conserved high levels of methylation were found in the CG context, 34% in the CHG context and 14.1% in the CHH context. The median levels of methylation at conserved sites were very high in all contexts within annotated genes and repetitive elements, and very low in genomic regions not known to contain genes or repetitive regions (Figure 15). There were 1,388 positions which were

100% methylated in all biological replicates of both *L. sativa* and *L. serriola* with an average read coverage of 24 reads and minimum read coverage of 11 reads. Forty-seven of these positions were found in seven genes, one of which was FLAGELLIN-SENSITIVE 2 a main sensor of bacterial infection; the remaining six genes are of unknown function (Table 8).

## Discussion

### *Methylome characterization*

When comparing the global methylation percentages in each sequence context of *L. sativa* and *L. serriola* to those of other genomes, a general positive relationship between percent global methylation and genome size is apparent, as has been detected previously using HPLC [110]. Similar to a recent report [111], we found that the relationship between methylation levels and genome size is sequence context dependent. There is a strong positive correlation between genome size and the total proportion methylation in the CG and CHG contexts, but that CHH methylation is only weakly correlated with genome size. Similarly, methylation levels at CG and CHG sites are highly correlated with methylation levels at sites within the same context that are separated by up to 50,000 bp, while methylation levels at neighboring CHH sites are less strongly correlated and that correlation does not vary significantly with distance. Both the correlation of methylation with genome size and spatial autocorrelation differences between the CG/CHG and the CHH contexts follow logically from the different mechanisms that maintain DNA methylation. Methylation in the CHH context must be

maintained by continual *de novo* RdDM targeted via small 21/24 nt RNA, where as methylation in the CG and CHG contexts is stably maintained through DNA replication by highly processive DNA methyltransferases MET1 and CMT3, [28].

DNA methylation has been proposed to act as a means of introducing stochasticity and evolutionary advantage to organisms in highly variable environments [74]. The overwhelming relative variability of CHH methylation in *L. sativa*, highlights the potential importance of differences in RdDM to significant phenotypic differences between *L. sativa* and *L. serriola*. We identified 1,697 cytosines having significantly different variances in the methylation between *L. sativa* biological replicates and *L. serriola* biological replicates. The majority of these positions were found in the RdDM driven CHH context (Figure 17). The proportion of DVCs which were more variable in the domestic *L. sativa* was much greater in the CHH context (92%) than in the CG (55%) or CHG (77%) contexts (Figure 18). The balance between RdDM and active demethylation by ROS1 may be finely turned for quick activation of pathogen defense response [41,112], an important phenotypic difference between *L. sativa* and *L. serriola*. Additionally, differences in the fidelity or kinetics of methylation or demethylation in the CHH context may contribute to differences in pathogen response and resistance. In Arabidopsis, loss of function mutants in the RdDM pathway harbored lower titers of bacteria than wild type plants [41]. If the RdDM pathway had a similar affect in *Lactuca* spp. this could contribute to the differences in pathogen tolerance between *L. serriola* relative to *L. sativa*.

The increased variation in methylation in domesticated lettuce is consistent with the increased methylation diversity seen in domesticated, relative to wild, soybeans [113]. These findings are particularly interesting in relation to the recent study by Latzel et al. that found populations of epigenetically diverse plants were more competitive as measured by increased plant density[75]. Increased planting density is a desirable trait selected for in agricultural production to increase crop yield, and the increased variability seen in *L. sativa* may be an adaptation to domestication.

In a pairwise comparison of sites in *L. sativa* and *L. serriola*, we found 13,623 differentially methylated cytosines (DMCs), most of which (92.19%) were not associated with known sequence polymorphisms between the species. Previous studies comparing methylation between related plant species have found the majority of methylation variation was associated with sequence diversity [29,30,54,114], though cases of pure epialleles, methylation variants independent of sequence, have been identified in maize [115], soybean [116], and *Arabidopsis* [31,117]. We found that the distribution of known SNPs was not significantly correlated with the distribution of DMCs in *L. sativa* and *L. serriola* and that the majority of DMCs between *L. sativa* and *L. serriola* did not co-localize with SNPs. Like Rambani et al. (2015) we found that the majority of DMC in protein coding genes were in the CG or CHG contexts (Figure 8 A and B) [118]. As with DVC, we found that the majority of DMCs were located in annotated repetitive regions (Figure 9 A, B and C). Becker et al. (2011) found fewer DMC among related *Arabidopsis* lines in regions near transposable elements, regions that were also enriched in small interfering RNAs [32]. However, we did not see an appreciable difference in the relative

frequency of DMC in protein coding genes or their upstream regions based on their proximity to annotated repetitive elements. Becker's strains were derived from a common ancestor 30 generations prior, whereas *L. sativa* and *L. serriola* are separated by centuries of selective breeding. It is possible the relative abundance of DMC and DVC in these regions in *Lactuca* species reflects an altered balance between RdDM and active demethylation between the species, and that the relatively short period of divergence between Becker's strains has not appreciably altered the balance between RdDM and active demethylation at these regions.

The majority of DMCs between *L. sativa* and *L. serriola* were more highly methylated in the domestic variety. Eichten et al. (2013) also found a positive association between location of differential methylation and repetitive elements, however they found that 81% of differentially methylated regions were more highly methylated in wild progenitor teosinte than in domesticated relative maize [114]. General categories of genes associated with domestication of crops, including transcription factors, enzymes and transporter proteins [119], were enriched among *Lactuca* genes containing a DMC.

In *Lactuca* both DMC and DVC are most prevalent in repetitive elements and could be associated with the high production of "sports" or rare unusual phenotypes in some varieties of *L. sativa* [120]. Though these sports appear to be generated randomly the variants, and their phenotypes, can be inherited similar to the stochastic generation [121,122] and inheritance with fidelity of differential methylation observed in offspring of regenerated plants [121]. Genomic stress of clonal propagation and tissue culture of plants are known to produce somatic variants or sports which are associated with

differential DNA methylation [121,122] and activity of transposable elements [123]. Though not grown clonally, the long history of inbreeding and selection to derive, in particular the heading varieties of, *L. sativa* from *L. serriola* could be thought of as a genomic stress.

It is tempting to hypothesize that the relative hypomethylation of DMCs in *L. serriola* may be associated with *L. serriola*'s superior performance in stressful and disturbed environments as has been suggested for other plant species. Resistance to biotic and non-biotic stresses have been associated with global [40,45,84] and loci specific [69] hypomethylation. Previous work in *A. thaliana* has shown that global loss of methylation is associated with increased resistance to infection by the bacterial pathogen *Pseudomonas syringae* [40]. Further, treatment of rice with the methyltransferase inhibitor 5-azadeoxycytidine induced global hypomethylation and resistance to infection by *Xanthomonas* [84]. In addition to the relative hypomethylation of *L. serriola* at DMC in general, all of the DMC within resistance genes (n=9) or within 1 kb upstream (n=10) or downstream (n=3) of resistance genes had higher percent methylation in *L. sativa* than *L. serriola* suggesting a particular role for methylation in and around these genes.

An interesting direction for future research would be to investigate active demethylation around resistance genes in *L. sativa* and *L. serriola*. The *L. sativa* and *L. serriola* genomes contain possible homologs to *A. thaliana* DNA glycosylases DEMETER and ROS1, but we did not find likely homologs to DML2 and DML3. In *A. thaliana* ROS1 is expressed in vegetative tissues while DEMETER is expressed in the endosperm and central cell during gametogenesis [124,125]. In *A. thaliana*, the region

between a Helitron transposon and ROS1's 5' UTR contains a DNA methylation monitoring sequence that is targeted by both RNA-directed DNA methylation and active de-methylation by ROS1 [126]. ROS1 expression is increased when methylated and decreased when de-methylated [52]. The active demethylation by ROS1 around genes in close proximity to repetitive elements has been associated with resistance to fungal and bacterial pathogens in *A. thaliana*. A triple demethylase mutant, deficient in all three known *A. thaliana* DNA glycosylases, was more susceptible to fungal infection and showed increased methylation and decreased expression of stress response genes with promoters in close proximity to transposons [69]. Yu et al. (2013) showed that growth of bacterial pathogen *Pseudomonas syringae* pv. tomato strain DC3000 was enhanced in ROS1 loss of function mutant, but not in loss of function mutants for the other DNA glycosylases, Demeter-like 2 (dml2) and Demeter-like 3 (dml3) [41]. Yu's work also showed that sRNAs accumulated in the region upstream of select resistance genes in Arabidopsis; regions where we see a spike in methylation in the CHH context in both species of *Lactuca*. Though the *L. sativa* and *L. serriola* methylation profiles around resistance genes are not significantly different on average from each other, analysis of differential rates of methylation and demethylation of resistance genes and flanking regions could highlight differences in pathogen responsiveness which could complement traditional gene based breeding approaches.

Our findings highlight significant differences in the methylomes of *L. sativa* and *L. serriola* that suggest future work in investigating epigenetic underpinnings of these closely related organisms' differing ability to adapt to disturbed environments.

## Figures

**Figure 1.** Genome-wide levels of methylation are highly reproducible between biological replicates of *L. sativa* and *L. serriola*. Genome-wide levels of methylation in the CG, CHG and CHHs contexts are highly consistent among biological replicates of *L. sativa* (A) and biological replicates of *L. serriola* (B).

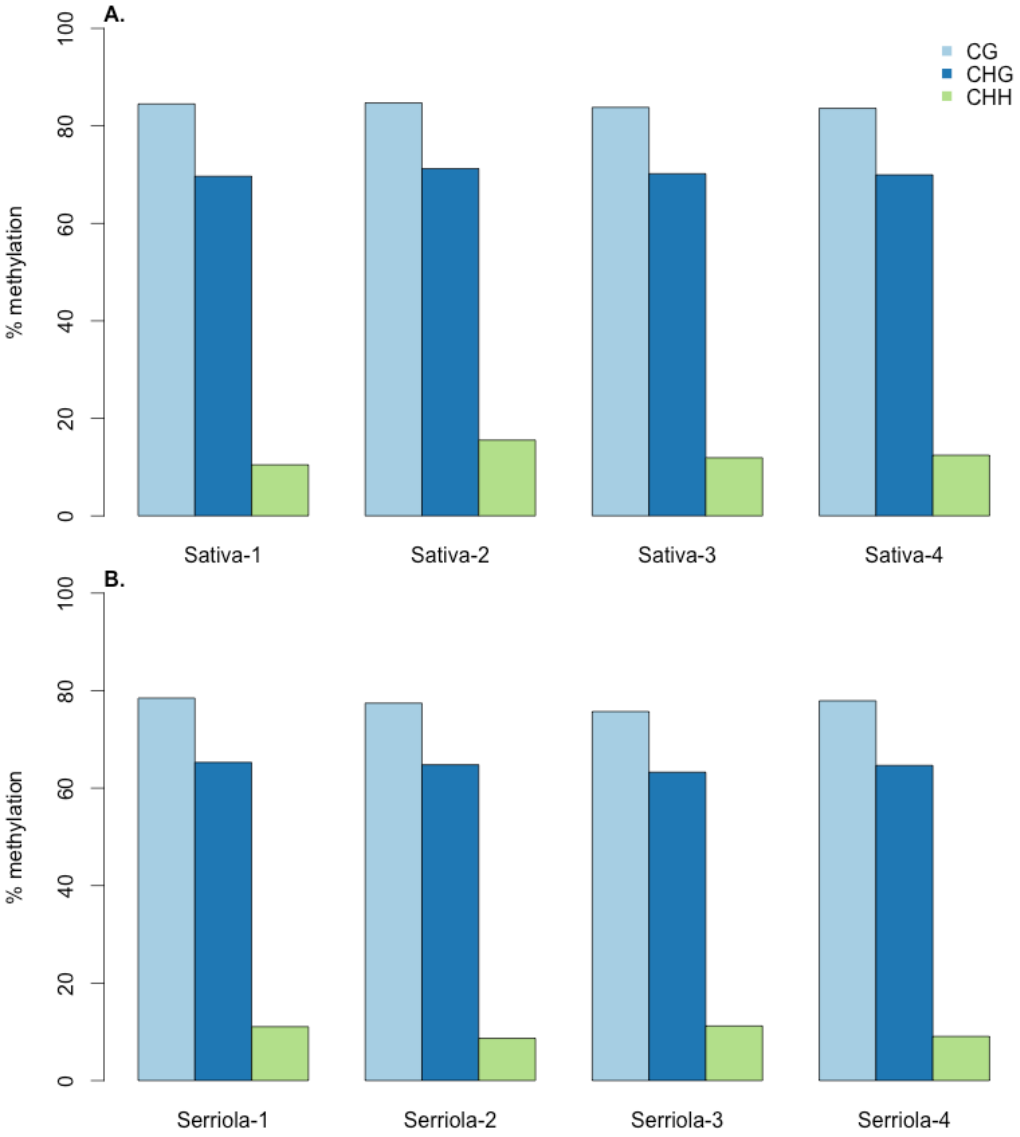


**Figure 2.** Genome-wide levels of methylation in plant species. Average methylation levels of *L. sativa* and *L. serriola* are high when compared to other plant methylomes especially in the CHH context (Figure 2A). Plants are listed from left to right in order of increasing genome size from *Arabidopsis thaliana* (135 MB) to *Zea mays* (2.5 GB), and literature reported genome sizes of each plant are plotted in Figure 2B. *Arabidopsis thaliana* [127], *Oryza sativa japonica* [128], *Oryza rufipogon* and *Oryza nivara* [129], *Oryza sativa indica* [130], *Brassica oleracea* [131], *Solanum lycopersicum* [132], *Glycine max* [133], *Zea maize* [134].

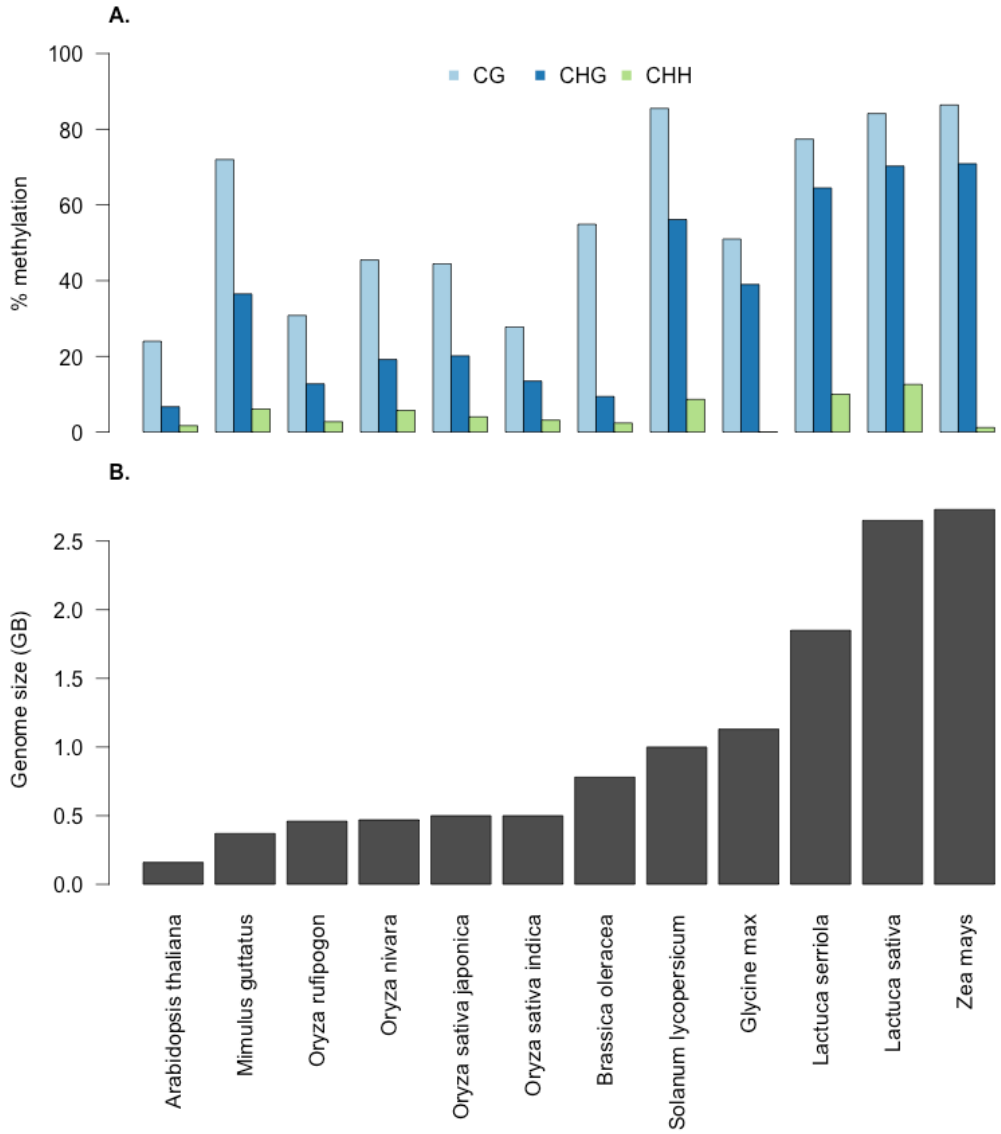


**Figure 3.** Correlation between genome size and genome-wide levels of methylation. There is a strong positive correlation between genome size and genome-wide levels of methylation in the CG (A) and CHG (B) sequence contexts for the 11 plant species shown in Figure 2. Methylation levels in the CHH sequence context (C) tend to be much lower and are not significantly correlated with genome size. Genome sizes are represented as C-values, the amount in picograms (pg) of DNA in a single copy of the genome.

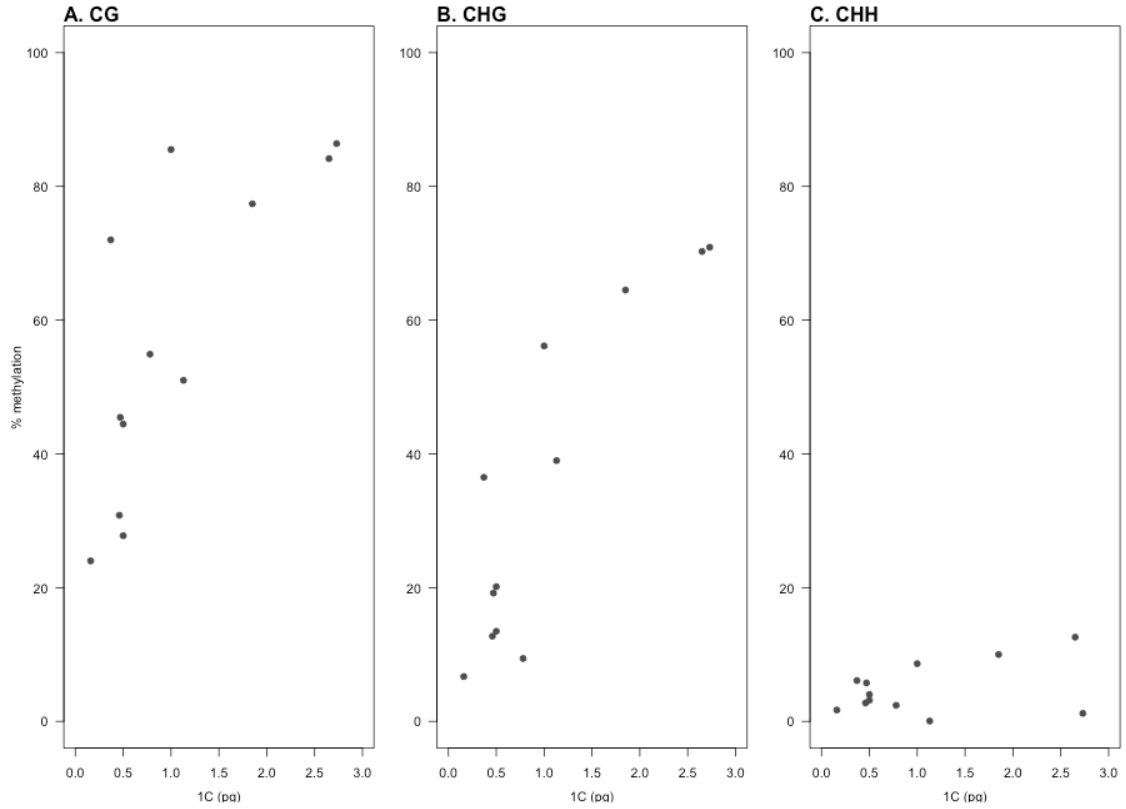


**Figure 4.** Inverse relationship of variation and average methylation levels. In all sequence contexts the coefficient of variance ("CV") of a position across biological replicates is inversely related to the percent methylation at that position in both *L. sativa* (A-C) and *L. serriola* (D-F) for cytosines covered by more than 10 reads in all samples.

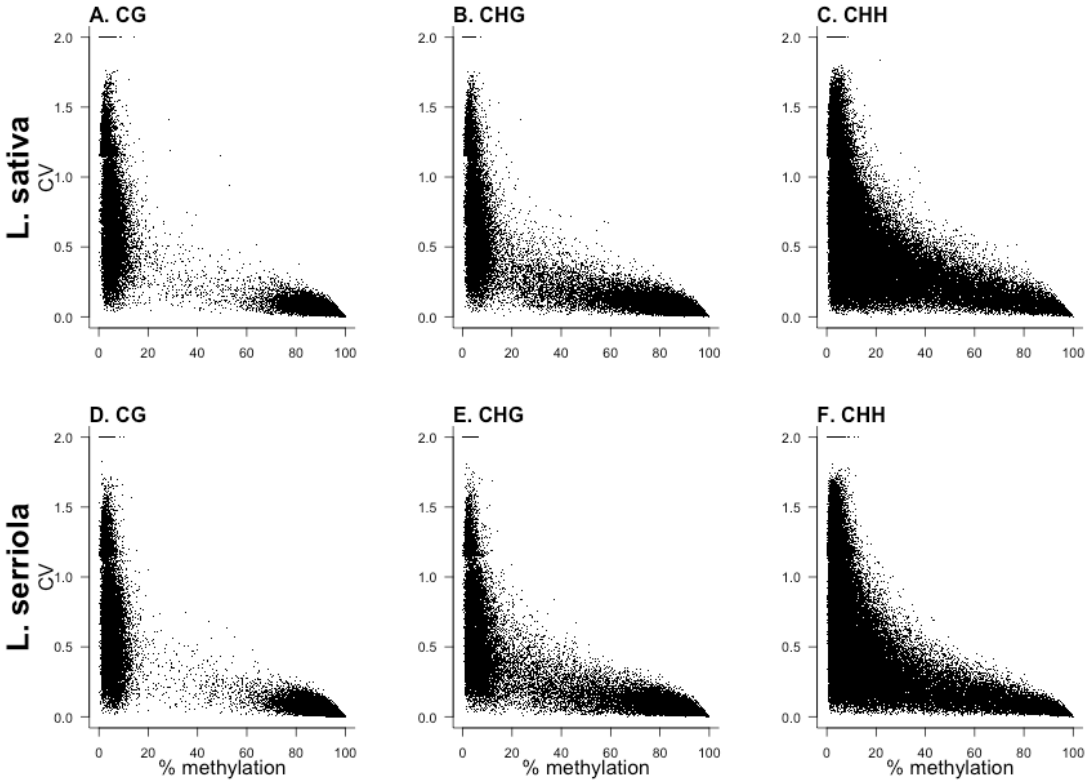


**Figure 5.** Spatial autocorrelation of methylation over long genomic distances. Methylation levels of cytosines in the CG (A) and CHG (B) contexts are highly correlated with methylation levels of cytosines in the same sequence context across increasing genomic distances, the x-axis refers to distance between adjacent positions and shows the correlation of positions separated by 0 to 100,000 bp. Cytosines in the CHH context (C) are not strongly correlated and this relationship is not affected by increasing genomic distance.

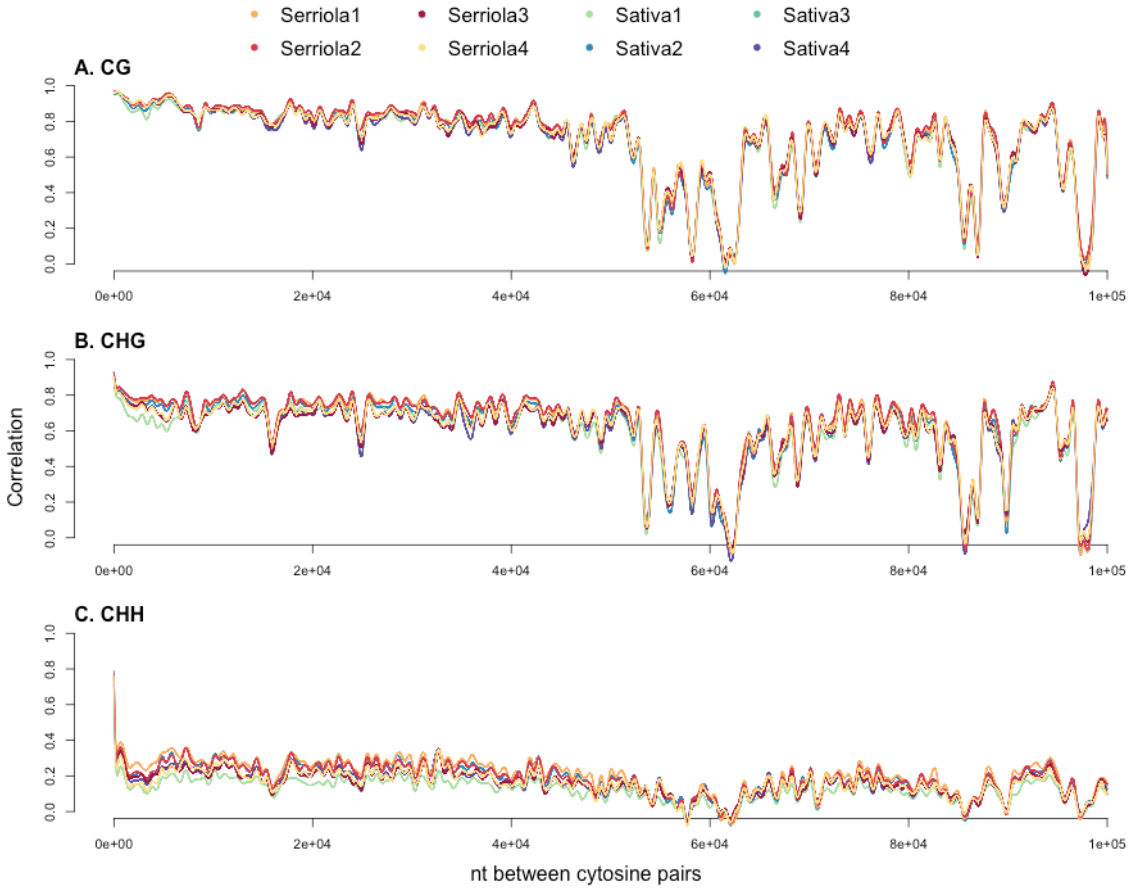


**Figure 6.** Spatial autocorrelation of methylation over short genomic distances. Methylation levels of cytosines in the CG (A) and CHG (B) contexts are highly correlated with methylation levels of cytosines in the same sequence context across increasing genomic distances, the x-axis refers to distance between adjacent positions and shows the correlation of positions separated by 0 to 1,000 bp. Cytosines in the CHH context (C) are strongly correlated only over very short (<100 bp) genomic regions, beyond this the degree of correlation is not affected by increasing genomic distance.

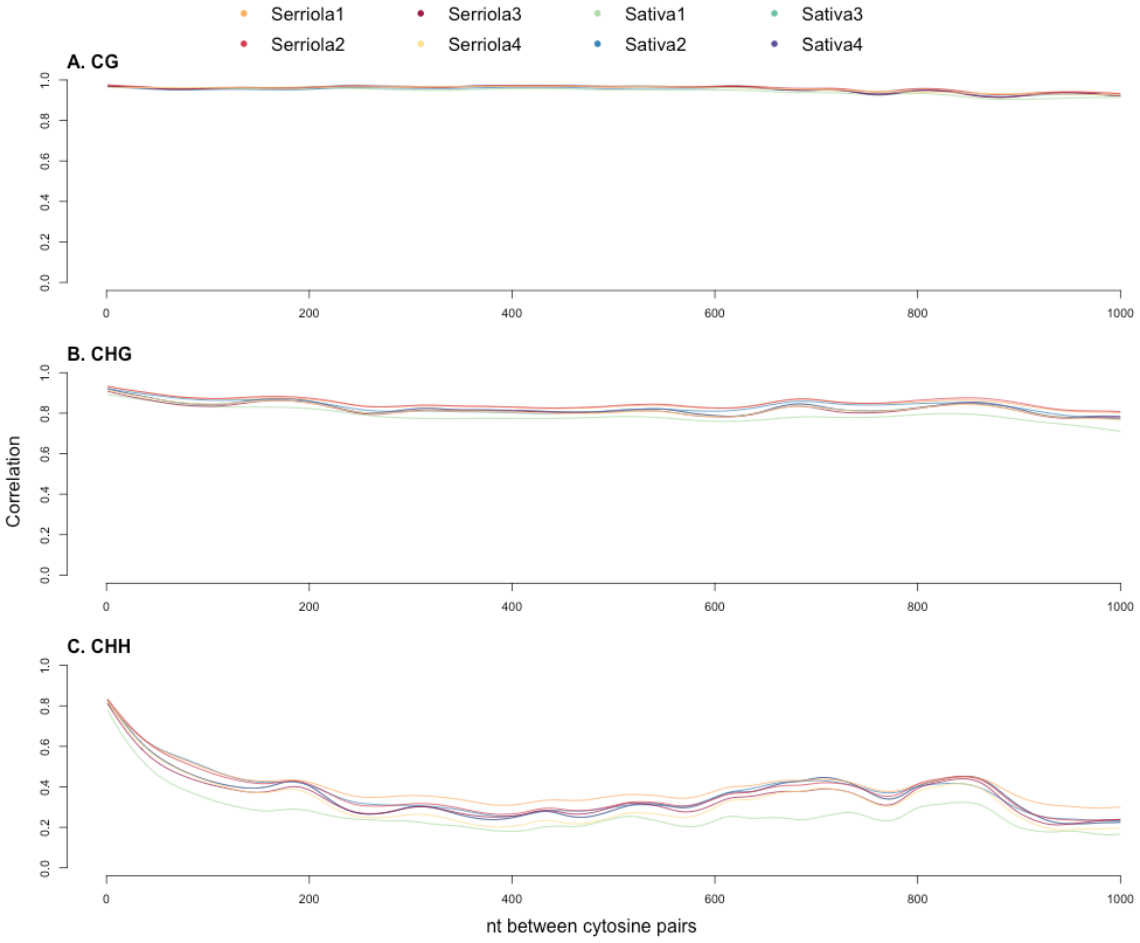


**Figure 7.** Average levels of methylation across protein coding genes and flanking regions. Average methylation in *L. sativa* for 10,000 bp preceding transcription start site, (TSS), average methylation over gene bodies where methylation averaged for each 100th of genes at least 1,000 bp long, and average methylation for 10,000 bp down stream of poly-A signal in CG (A), CHG (B) and CHH (C) contexts. *L. sativa* and *L. serriola* results do not differ appreciably, *L. serriola* results not shown.

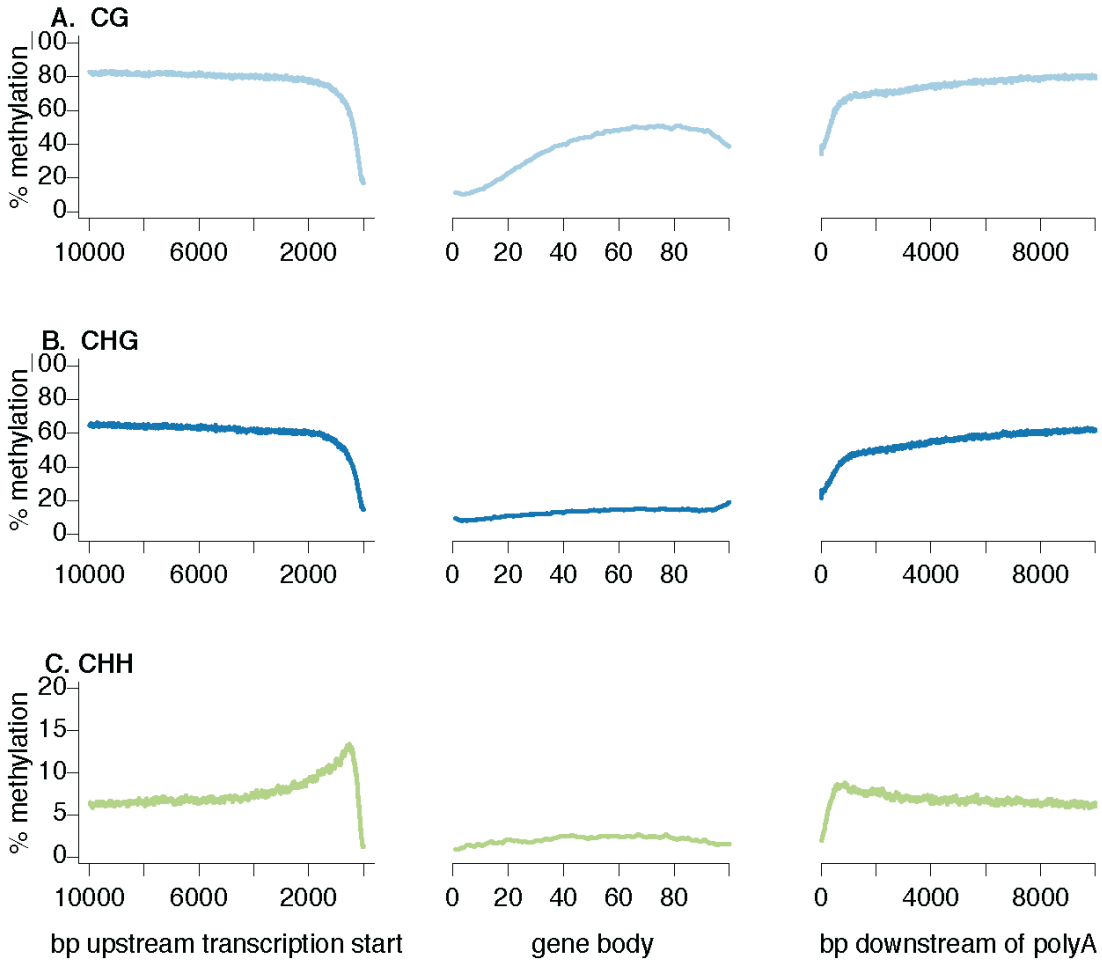


**Figure 8.** Average levels of methylation across resistance genes and flanking regions. Average methylation of resistance genes in *L. sativa* for 1,000 bp preceding transcription start site, average methylation over gene bodies where methylation averaged for each 100th of genes at least 1,000 bp long, and average methylation for 1,000 bp down stream of poly-A signal in CH (A), CHG (B) and CHH (C) contexts. *L. sativa* and *L. serriola* results do not differ appreciably, *L. serriola* results not shown.

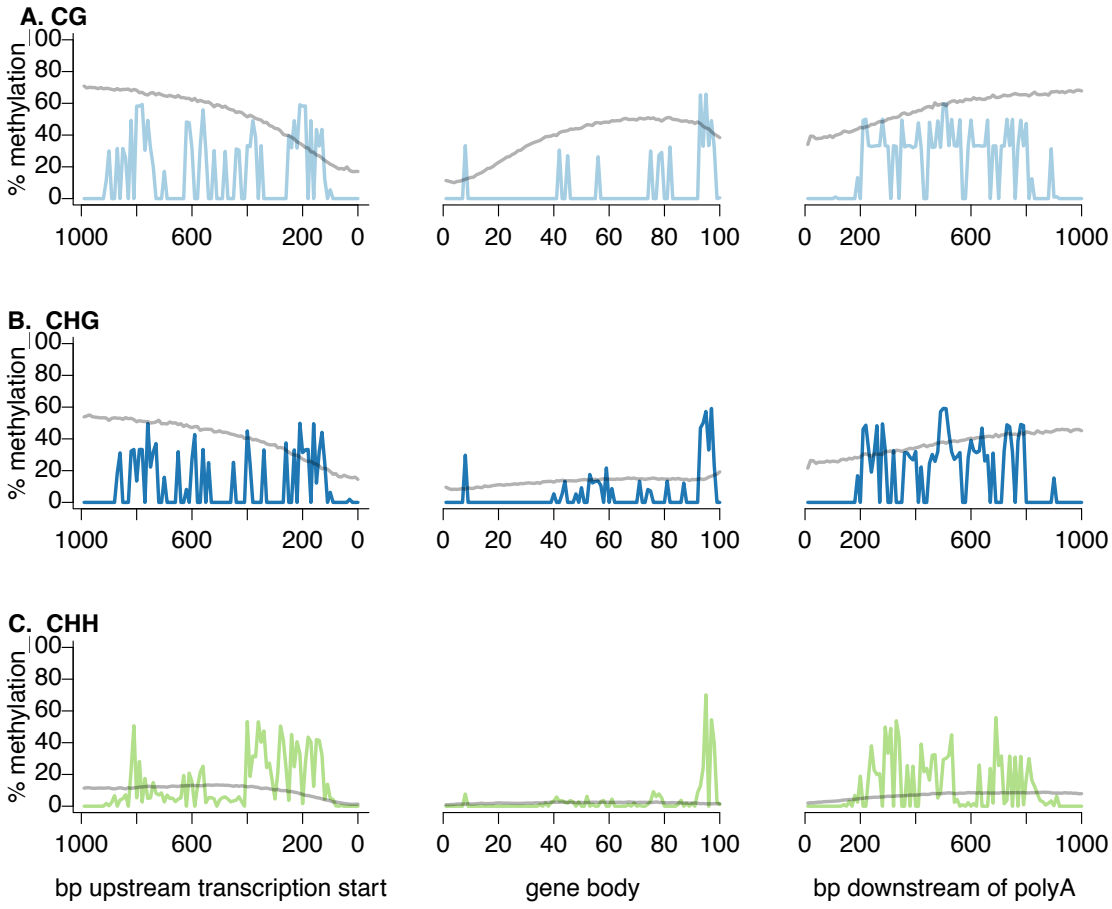


**Figure 9**: Distribution of DMCs across genomic regions. The majority of DMCs are located in repetitive regions in CG (A), CHG (B), and CHH (C) sequence contexts. The majority of DMCs in repetitive regions are located in CG (D), CHG (E), and CHH (F) sequence contexts in long terminal repeat (LTR) retrotransposons families.

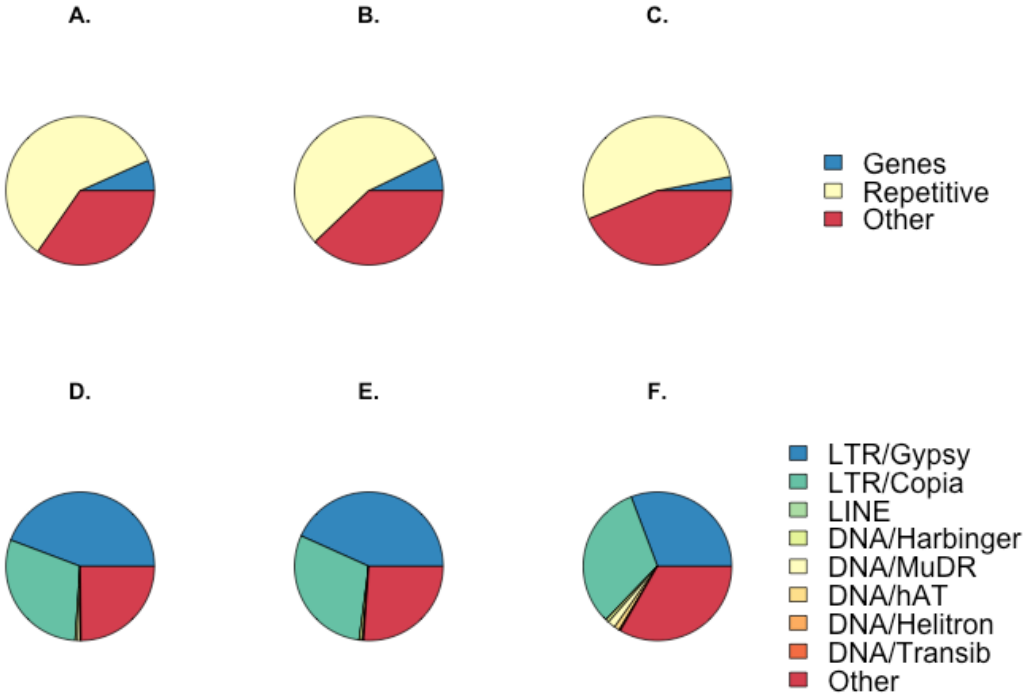


**Figure 10.** Relative number of DMCs by sequence context and proximity of feature to annotated repetitive regions. The number of DMCs per protein coding gene (A.) or upstream region (B.) does not differ appreciably by proximity to repetitive regions.

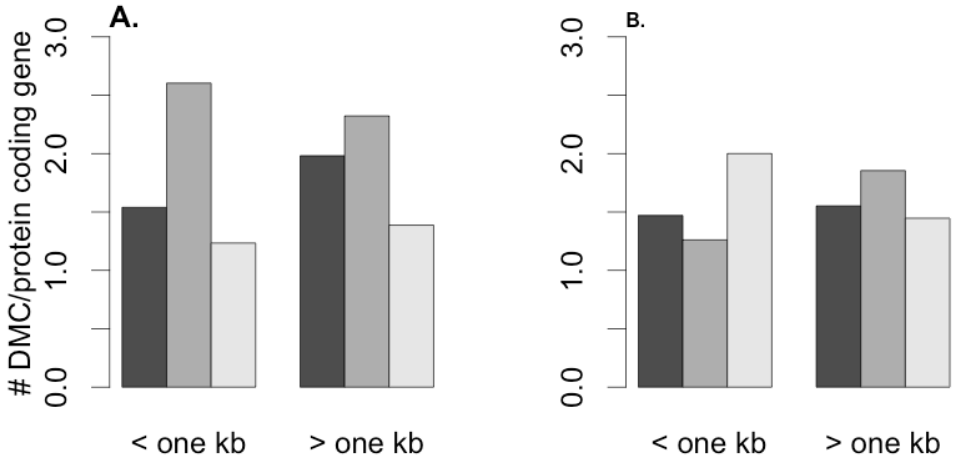


**Figure 11.** Percent methylation of DMCs in *L. sativa* and *L. serriola* by sequence context and proximity of feature to annotated repetitive regions. Percentage methylation of DMCs by sequence context and genotype for protein coding genes within 1 kb of repeats (A) and not within 1 kb of repeats (B); for upstream regions of protein coding genes within 1 kb of repeats (C) and upstream regions not within 1 kb of repeats (D) Error bars represent 95% confidence interval.

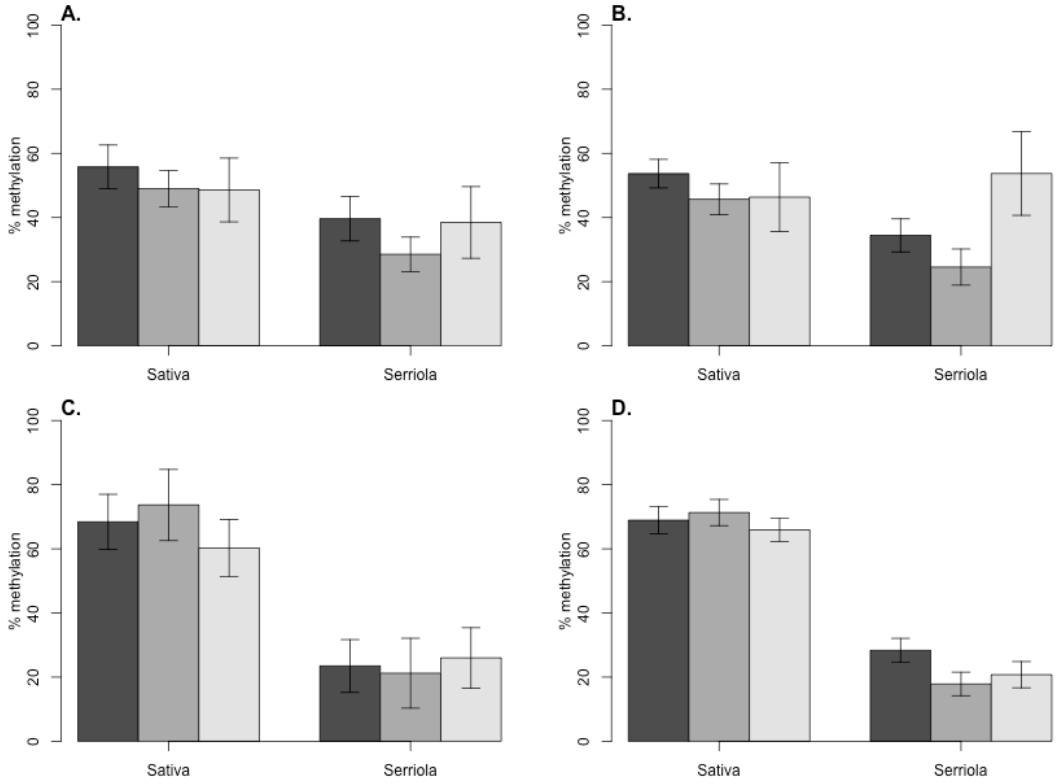


**Figure 12.** Distribution of DVCs by genomic region. The majority of differentially variable cytosines are located in repetitive regions in CG (A), CHG (B), and CHH (C) sequence contexts. The majority of differentially variable cytosines in repetitive regions are located in CG (D), CHG (E), and CHH (F) sequence contexts in long terminal repeat (LTR) retrotransposons families.

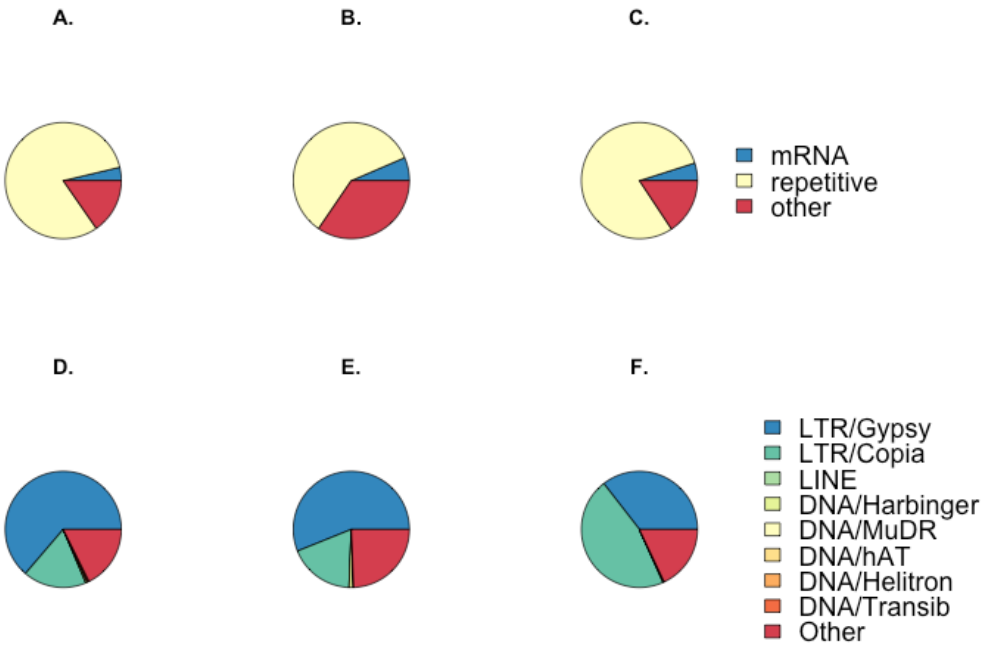


**Figure 13.** The locations of DVCs and DMCs showed low to moderate correlation of abundance across the genome. Spearman correlation of the frequency of protein coding genes (mrna_counts) and repetitive elements (repeat_counts) and single nucleotide polymorphisms (snp_counts) with all differentially methylated cytosines combined (all_mc_counts) and all differentially variable cytosines combined (all_var_counts) (B) or differentially methylated cytosines by sequence context (cg_mc_counts, chg_mc_counts, chh_mc_counts) and differentially variable cytosines by sequence context (cg_var_counts, chg_var_counts, chh_var_counts) (A).

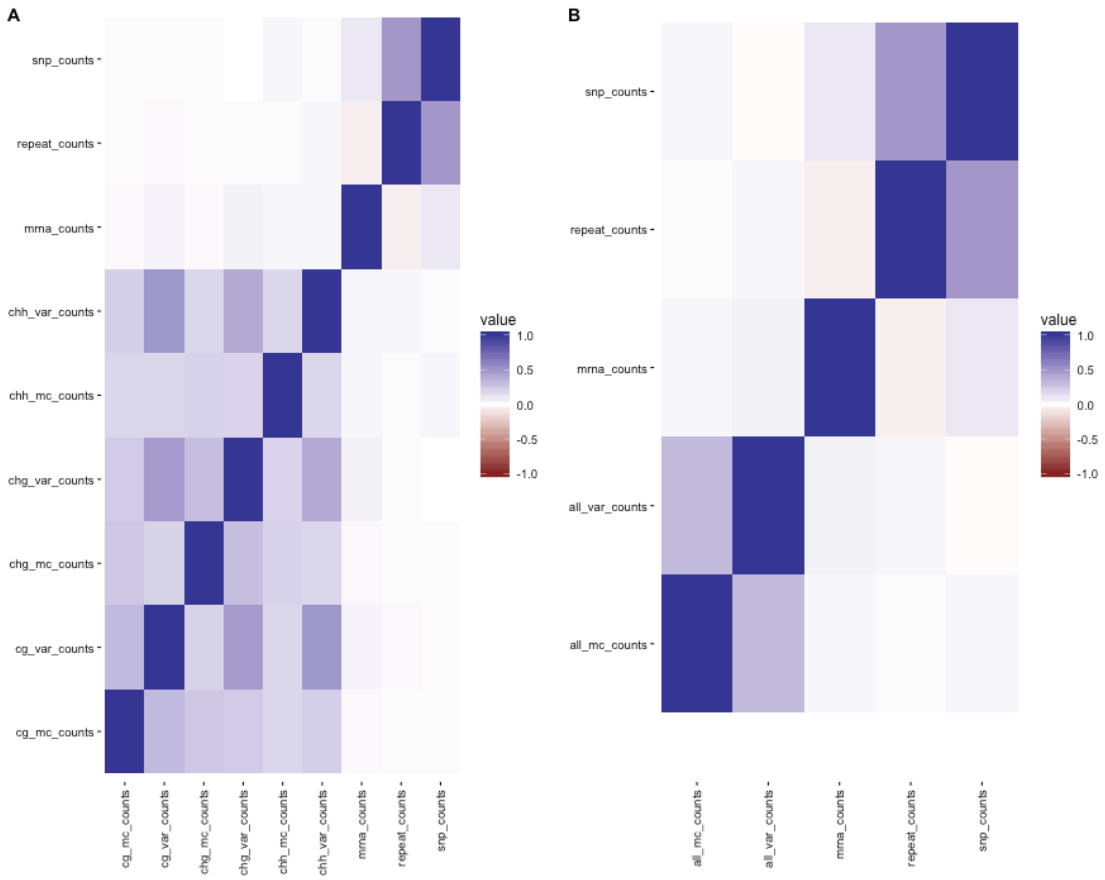


**Figure 14.** Distribution of sites of conserved methylation by genomic region. Distribution of conserved methylated cytosines between *L. sativa* and *L. serriola* in CG (A), CHG (B), and CHH (C) sequence contexts in gene bodies, repetitive elements or other genomic regions, not known to contain genes or repetitive regions.

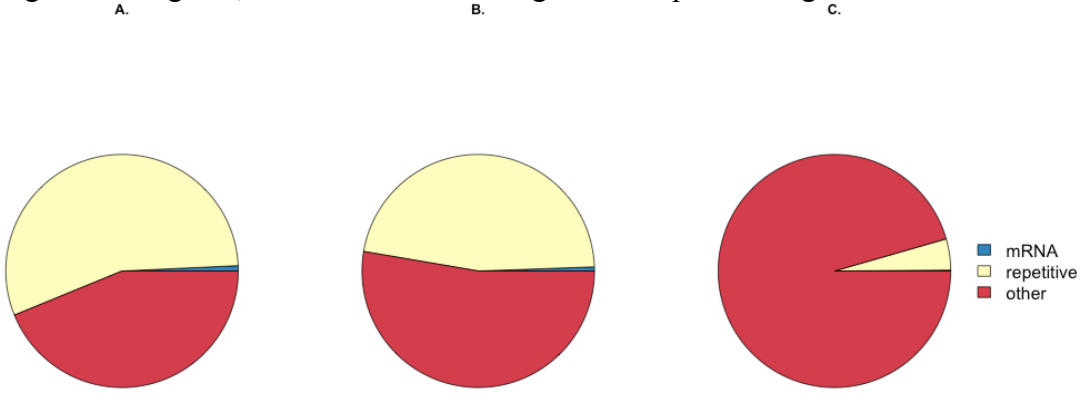


**Figure 15.** Percent methylation at conserved sites by genomic region. Sites of conserved methylation found within genes or repetitive regions have extremely high levels of methylation, whereas very low levels of methylation are found at sites of conversed methylation found in other genomic regions, not known to contain genes or repetitive regions.

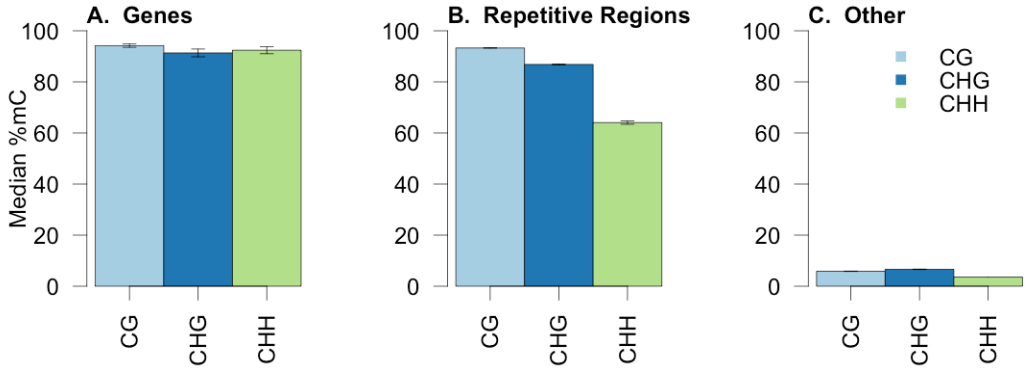


**Figure 16.** Frequency of sites of conserved methylation by average methylation level. The majority of positions with highly conserved methylation states were found at positions which had very low levels of methylation. The graph shows the frequency of conserved positions arranged in order of increasing average methylation level. Positions with highly conserved methylation levels are those positions where the variability of percent methylation between biological replicates of each genotype was in the lowest 25% for that genotype and the methylation percentage between biological replicates of *L. sativa* and biological replicates of *L. serriola* differed by less than 20%.

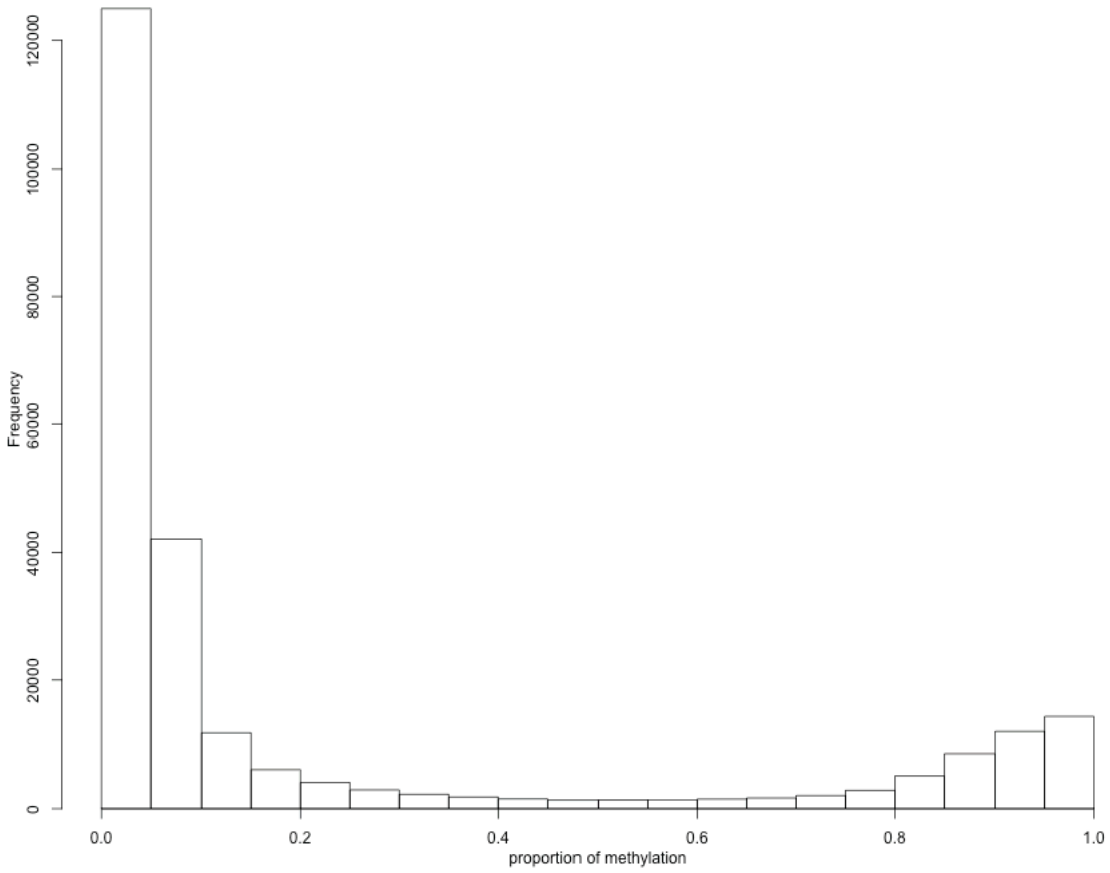


**Figure 17.** Frequency of DVCs by sequence context. The majority of DVCs between *L. sativa* and *L. serriola* are found in the CHH context.

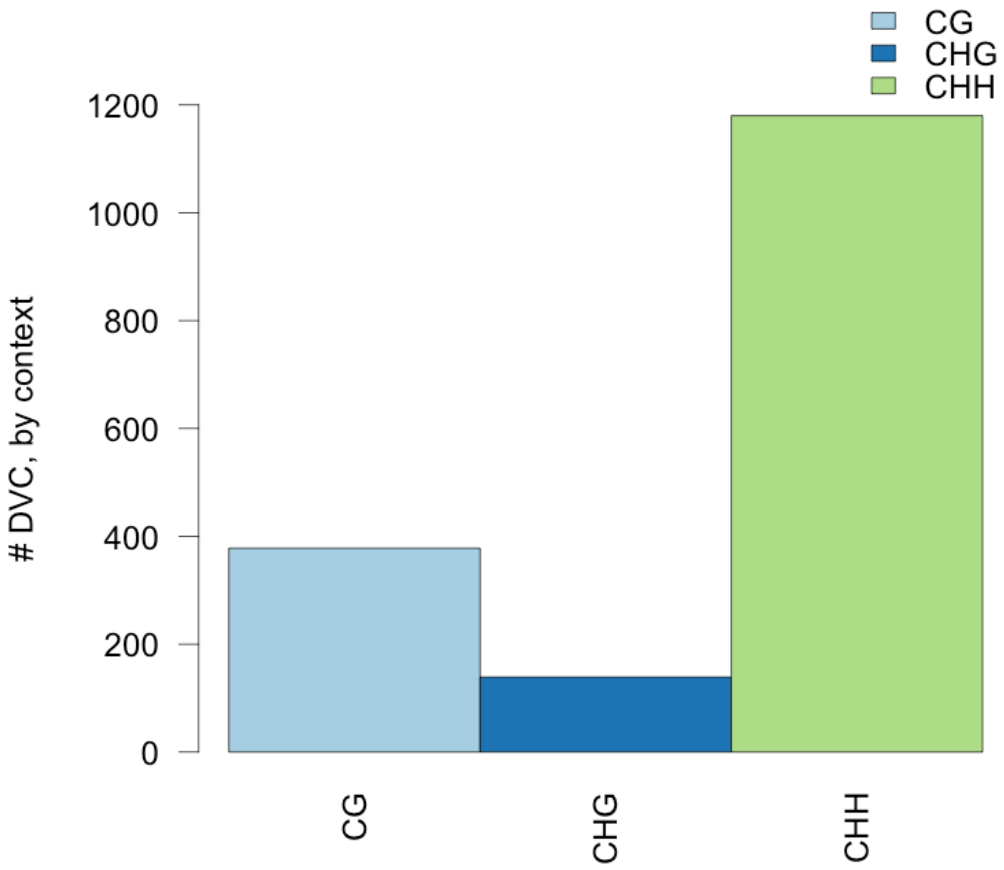
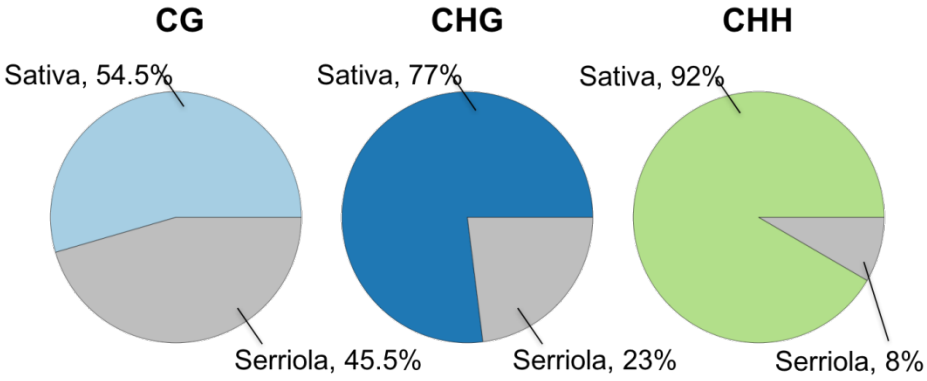


**Figure 18.** Proportion of DVCs that are more variable in *L. sativa* or *L. serriola*. Most differentially variable cytosines have higher levels of variability in *L. sativa* than in *L. serriola*.



## Tables

**Table 1.** Protein coding genes with frequent occurrence of DMCs. The location column indicates whether or not the gene is located within 1 kb of an annotated repetitive element.

| mRNA ID | Location | Freq. DMC | Average % mC *L. sativa* | Average % mC *L. serriola* | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|---|
| Lsat_1_v5_gn_5_157381 | one kb | 53 | 42.52687331 | 16.65685226 | GO:0004553: hydrolase activity, hydrolyzing O-glycosyl compounds: Molecular Function | endo-1,4-beta-mannosidase (EC:3.-.-.-) \| K01567 [EC:3.-.-.-] |
| Lsat_1_v5_gn_3_81720 | more than one kb | 23 | 51.26227875 | 2.423321597 | | |
| Lsat_1_v5_gn_4_121121 | more than one kb | 18 | 40.47586824 | 9.118170134 | GO:0004497: monooxygenase activity: Molecular Function \| GO:0005506: iron ion binding: Molecular Function \| GO:0009055: electron carrier activity: Molecular Function \| GO:0020037: heme binding: Molecular Function | CYP82C4 \| electron carrier/ heme binding / iron ion binding / monooxygenase/ oxygen binding \| K00517 [EC:1.14.-.-] |
| Lsat_1_v5_gn_5_60241 | one kb | 18 | 86.64087226 | 4.058201301 | | |
| Lsat_1_v5_gn_4_120360 | more than one kb | 17 | 32.75042683 | 61.64468393 | GO:0004497: monooxygenase activity: Molecular Function \| GO:0005506: iron ion binding: Molecular Function \| GO:0009055: electron carrier activity: Molecular Function \| GO:0020037: heme binding: Molecular Function | CYP82C4 \| electron carrier/ heme binding / iron ion binding / monooxygenase/ oxygen binding \| K00517 [EC:1.14.-.-] |
| Lsat_1_v5_gn_4_186001 | more than one kb | 15 | 46.26032561 | 87.42532016 | | |
| Lsat_1_v5_gn_8_52201 | more than one kb | 14 | 35.47441458 | 8.743896814 | | EFR \| EFR (EF-TU RECEPTOR) \| ATP binding / kinase/ protein |

| mRNA ID | Location | Freq. DMC | Average % mC *L. sativa* | Average % mC *L. serriola* | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|---|
| | | | | | | serine/threonine kinase \| K13428 LRR receptor-like serine/threonine-protein kinase EFR [EC:2.7.11.1] |
| Lsat_1_v5_gn_8_51501 | more than one kb | 12 | 44.96402592 | 11.99408829 | GO:0004888: transmembrane receptor activity: Molecular Function \| GO:0005524: ATP binding: Molecular Function \| GO:0006915: apoptosis: Biological Process \| GO:0007165: signal transduction: Biological Process \| GO:0031224: intrinsic to membrane: Cellular Component | |
| Lsat_1_v5_gn_8_51520 | more than one kb | 11 | 44.04897287 | 1.09796669 | | EFR \| EFR (EF-TU RECEPTOR) \| ATP binding / kinase/ protein serine/threonine kinase \| K13428 LRR receptor-like serine/threonine-protein kinase EFR [EC:2.7.11.1] |
| Lsat_1_v5_gn_7_103821 | more than one kb | 10 | 25.11171772 | 72.65490565 | | hypothetical protein LOC100252654 \| K10302 F-box protein 22 |
| Lsat_1_v5_gn_1_116300 | more than one kb | 9 | 51.25746895 | 1.266484043 | | |
| Lsat_1_v5_gn_2_37560 | more than one kb | 9 | 3.489673888 | 86.27979218 | GO:0016020: membrane: Cellular Component \| GO:0030001: metal ion transport: Biological Process \| GO:0046873: metal ion transmembrane transporter activity: Molecular Function | |

| mRNA ID | Location | Freq. DMC | Average % mC *L. sativa* | Average % mC *L. serriola* | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|---|
| Lsat_1_v5_gn_2_79801 | more than one kb | 8 | 70.64564469 | 1.383656089 | GO:0004672: protein kinase activity: Molecular Function \| GO:0004674: protein serine/threonine kinase activity: Molecular Function \| GO:0005488: binding: Molecular Function \| GO:0005524: ATP binding: Molecular Function | protein kinase, putative \| K00924 [EC:2.7.1.-] |
| Lsat_1_v5_gn_2_131781 | more than one kb | 8 | 62.66904651 | 22.30709298 | | |
| Lsat_1_v5_gn_4_118880 | one kb | 7 | 30.52014604 | 0.831021859 | GO:0004252: serine-type endopeptidase activity: Molecular Function | clpP \| ATP-dependent Clp protease proteolytic subunit \| K01358 ATP-dependent Clp protease, protease subunit [EC:3.4.21.92] |
| Lsat_1_v5_gn_7_15020 | one kb | 7 | 9.758959529 | 90.23934699 | GO:0003677: DNA binding: Molecular Function \| GO:0003700: sequence-specific DNA binding transcription factor activity: Molecular Function \| GO:0005515: protein binding: Molecular Function \| GO:0005634: nucleus: Cellular Component \| GO:0006355: regulation of transcription, DNA-dependent: Biological Process | Pbx4, Edg4 \| pre-B-cell leukemia homeobox 4 \| K09355 pre-B-cell leukemia transcription factor |
| Lsat_1_v5_gn_4_75780 | one kb | 7 | 1.69929683 | 86.35457316 | | pex7 \| WD40 repeat-containing protein \| K13341 peroxin-7 |
| Lsat_1_v5_gn_7_95881 | more than one kb | 7 | 43.32149422 | 3.53940139 | GO:0003677: DNA binding: Molecular Function \| | rpoB \| RNA polymerase beta subunit (EC:2.7.7.6) \| K03043 DNA- |

| mRNA ID | Location | Freq. DMC | Average % mC *L. sativa* | Average % mC *L. serriola* | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|---|
| | | | | | GO:0003899: DNA-directed RNA polymerase activity: Molecular Function | directed RNA polymerase subunit beta [EC:2.7.7.6] |
| Lsat_1_v5_gn_1_49901 | more than one kb | 6 | 0.757508579 | 93.33249082 | | hypothetical protein LOC100247694 \| K11583 protein phosphatase 2 (formerly 2A), regulatory subunit B"Scaffold=Lsat_1_v5_g_1_2399 |
| Lsat_1_v5_gn_2_34800 | more than one kb | 6 | 54.35380201 | 6.387451436 | GO:0009521: photosystem: Cellular Component \| GO:0009767: photosynthetic electron transport chain: Biological Process \| GO:0016020: membrane: Cellular Component \| GO:0016168: chlorophyll binding: Molecular Function | psbC \| photosystem II 44 kDa protein \| K02705 photosystem II CP43 chlorophyll apoprotein |
| Lsat_1_v5_gn_5_36381 | one kb | 6 | 0 | 94.0648275 | | |
| Lsat_1_v5_gn_7_7020 | one kb | 6 | 10.02364161 | 64.59990596 | | |
| Lsat_1_v5_gn_4_120300 | more than one kb | 5 | 39.69122293 | 5.909350423 | GO:0004497: monooxygenase activity: Molecular Function \| GO:0005506: iron ion binding: Molecular Function \| GO:0009055: electron carrier activity: Molecular Function \| GO:0020037: heme binding: Molecular Function | CYP82C4 \| electron carrier/ heme binding / iron ion binding / monooxygenase/ oxygen binding \| K00517 [EC:1.14.-.-] |
| Lsat_1_v5_gn_1_10481 | more than one kb | 5 | 1.670997948 | 60.01399436 | GO:0003677: DNA binding: Molecular Function \| GO:0003824: catalytic | DNA polymerase I (POL I) \| K02335 DNA polymerase I [EC:2.7.7.7] |

| mRNA ID | Location | Freq. DMC | Average % mC *L. sativa* | Average % mC *L. serriola* | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|---|
| | | | | | activity: Molecular Function | |
| Lsat_1_v5_gn_4_144541 | one kb | 5 | 98.54095195 | 10 | | |
| Lsat_1_v5_gn_4_183940 | more than one kb | 5 | 71.83309789 | 46.59597312 | | |
| Lsat_1_v5_gn_6_37880 | more than one kb | 5 | 49.46149523 | 3.544937009 | | |
| Lsat_1_v5_gn_8_102340 | more than one kb | 5 | 97.74642394 | 41.81293883 | | |
| Lsat_1_v5_gn_1_8140 | more than one kb | 4 | 87.70722568 | 4.131184885 | GO:0004497: monooxygenase activity: Molecular Function \| GO:0005506: iron ion binding: Molecular Function \| GO:0009055: electron carrier activity: Molecular Function \| GO:0020037: heme binding: Molecular Function | CYP81D3 \| electron carrier/ heme binding / iron ion binding / monooxygenase/ oxygen binding \| K00517 [EC:1.14.-.-] |
| Lsat_1_v5_gn_4_23500 | more than one kb | 4 | 86.26581025 | 13.99313738 | | hypothetical protein \| K10610 DNA damage-binding protein 1 |
| Lsat_1_v5_gn_4_154701 | more than one kb | 4 | 82.49343656 | 0.823882901 | | hypothetical protein LOC100260814 \| K01873 valyl-tRNA synthetase [EC:6.1.1.9]Scaffold=Lsat_1_v5_g_4_137 |
| Lsat_1_v5_gn_3_92480 | one kb | 4 | 79.76345118 | 0 | GO:0004672: protein kinase activity: Molecular Function \| GO:0004674: protein serine/threonine kinase activity: Molecular Function \| GO:0005524: ATP binding: Molecular Function | SNRK2.2 (SNF1-RELATED PROTEIN KINASE 2.2) \| kinase/ protein kinase \| K00924 [EC:2.7.1.-] |
| Lsat_1_v5_gn_4_54820 | one kb | 4 | 6.34674667 | 98.43807557 | GO:0006629: lipid metabolic process: Biological Process | Zeta-carotene desaturase, chloroplast precursor, putative (EC:1.14.99.30) \| K00514 zeta-carotene desaturase |

| mRNA ID | Location | Freq. DMC | Average % mC *L. sativa* | Average % mC *L. serriola* | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|---|
| | | | | | | [EC:1.14.99.30] |
| Lsat_1_v5_gn_2_95981 | one kb | 4 | 95.73928962 | 0 | | |
| Lsat_1_v5_gn_3_17360 | more than one kb | 4 | 57.83039578 | 1.229783353 | | |
| Lsat_1_v5_gn_4_167780 | more than one kb | 4 | 41.02307881 | 70.2750932 | | |
| Lsat_1_v5_gn_7_67161 | one kb | 4 | 58.23682653 | 0.733750942 | | |
| Lsat_1_v5_gn_8_105980 | one kb | 4 | 73.80332287 | 10.27888622 | | |
| Lsat_1_v5_gn_8_125081 | more than one kb | 4 | 32.34516424 | 97.99682517 | | |
| Lsat_1_v5_gn_8_94660 | one kb | 4 | 48.62625699 | 47.15131292 | | |

**Table 2.** Enriched gene ontology terms for protein coding genes containing one or more DMCs.

| GO ID | GO description | Subset count | Genome count | Raw p-value | Adj. p-value |
|---|---|---|---|---|---|
| GO:0008289 | lipid binding: Molecular Function | 1 | 1 | 0 | 0 |
| GO:0051287 | NAD or NADH binding: Molecular Function | 2 | 3 | 1.1894E-06 | 7.43376E-05 |
| GO:0048038 | quinone binding: Molecular Function | 3 | 20 | 5.30646E-05 | 0.002211023 |
| GO:0009521 | photosystem: Cellular Component | 2 | 10 | 0.000135029 | 0.002813112 |
| GO:0009767 | photosynthetic electron transport chain: Biological Process | 2 | 10 | 0.000135029 | 0.002813112 |
| GO:0016168 | chlorophyll binding: Molecular Function | 2 | 10 | 0.000135029 | 0.002813112 |
| GO:0003855 | 3-dehydroquinate dehydratase activity: Molecular Function | 1 | 3 | 0.000335334 | 0.004657422 |
| GO:0006467 | protein thiol-disulfide exchange: Biological Process | 1 | 3 | 0.000335334 | 0.004657422 |
| GO:0008964 | phosphoenolpyruvate carboxylase activity: Molecular Function | 1 | 3 | 0.000335334 | 0.004657422 |
| GO:0006099 | tricarboxylic acid cycle: Biological Process | 1 | 4 | 0.000665949 | 0.006936966 |
| GO:0006334 | nucleosome assembly: Biological Process | 1 | 4 | 0.000665949 | 0.006936966 |
| GO:0008250 | oligosaccharyltransferase complex: Cellular Component | 1 | 4 | 0.000665949 | 0.006936966 |
| GO:0004556 | alpha-amylase activity: Molecular Function | 1 | 5 | 0.00110211 | 0.00918425 |
| GO:0004765 | shikimate kinase activity: Molecular Function | 1 | 5 | 0.00110211 | 0.00918425 |
| GO:0005789 | endoplasmic reticulum membrane: Cellular Component | 2 | 19 | 0.001015632 | 0.00918425 |
| GO:0016651 | oxidoreductase activity, acting on NADH or NADPH: Molecular Function | 2 | 20 | 0.001185497 | 0.009261695 |
| GO:0004252 | serine-type endopeptidase activity: Molecular Function | 5 | 115 | 0.0014697 | 0.010799674 |
| GO:0004579 | dolichyl-diphosphooligosaccharide-protein glycotransferase activity: Molecular Function | 1 | 6 | 0.001641551 | 0.010799674 |
| GO:0016671 | oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor: Molecular Function | 1 | 6 | 0.001641551 | 0.010799674 |
| GO:0004888 | transmembrane receptor activity: Molecular Function | 7 | 210 | 0.00192669 | 0.01146839 |
| GO:0031224 | intrinsic to membrane: Cellular Component | 7 | 209 | 0.001870177 | 0.01146839 |
| GO:0004764 | shikimate 5-dehydrogenase activity: Molecular Function | 1 | 7 | 0.002282039 | 0.01296613 |
| GO:0004659 | prenyltransferase activity: Molecular Function | 1 | 11 | 0.005811072 | 0.031581912 |
| GO:0016760 | cellulose synthase (UDP-forming) activity: Molecular Function | 2 | 36 | 0.006550426 | 0.034116801 |
| GO:0004519 | endonuclease activity: Molecular Function | 1 | 14 | 0.009414681 | 0.042029824 |
| GO:0009772 | photosynthetic electron transport in photosystem II: Biological Process | 1 | 14 | 0.009414681 | 0.042029824 |
| GO:0016853 | isomerase activity: Molecular Function | 1 | 14 | 0.009414681 | 0.042029824 |
| GO:0030077 | plasma membrane light-harvesting | 1 | 14 | 0.009414681 | 0.042029824 |

| GO ID | GO description | Subset count | Genome count | Raw p-value | Adj. p-value |
|---|---|---|---|---|---|
| | complex: Cellular Component | | | | |
| GO:0019684 | photosynthesis, light reaction: Biological Process | 1 | 15 | 0.010787375 | 0.044746869 |
| GO:0001522 | pseudouridine synthesis: Biological Process | 1 | 18 | 0.015392923 | 0.044746869 |
| GO:0004143 | diacylglycerol kinase activity: Molecular Function | 1 | 17 | 0.013778379 | 0.044746869 |
| GO:0005643 | nuclear pore: Cellular Component | 1 | 17 | 0.013778379 | 0.044746869 |
| GO:0006855 | drug transmembrane transport: Biological Process | 2 | 49 | 0.015278703 | 0.044746869 |
| GO:0006952 | defense response: Biological Process | 6 | 238 | 0.014121969 | 0.044746869 |
| GO:0008131 | primary amine oxidase activity: Molecular Function | 1 | 16 | 0.012242574 | 0.044746869 |
| GO:0008171 | O-methyltransferase activity: Molecular Function | 1 | 17 | 0.013778379 | 0.044746869 |
| GO:0009308 | amine metabolic process: Biological Process | 1 | 16 | 0.012242574 | 0.044746869 |
| GO:0009522 | photosystem I: Cellular Component | 1 | 18 | 0.015392923 | 0.044746869 |
| GO:0015238 | drug transmembrane transporter activity: Molecular Function | 2 | 49 | 0.015278703 | 0.044746869 |
| GO:0015297 | antiporter activity: Molecular Function | 2 | 49 | 0.015278703 | 0.044746869 |
| GO:0015977 | carbon fixation: Biological Process | 1 | 18 | 0.015392923 | 0.044746869 |
| GO:0016165 | lipoxygenase activity: Molecular Function | 1 | 17 | 0.013778379 | 0.044746869 |
| GO:0016702 | oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen: Molecular Function | 1 | 17 | 0.013778379 | 0.044746869 |
| GO:0005507 | copper ion binding: Molecular Function | 4 | 139 | 0.016669006 | 0.047355131 |
| GO:0009579 | thylakoid: Cellular Component | 1 | 19 | 0.017084369 | 0.04745658 |
| GO:0007165 | signal transduction: Biological Process | 7 | 308 | 0.01782667 | 0.048442037 |
| GO:0004629 | phospholipase C activity: Molecular Function | 1 | 20 | 0.01885091 | 0.049090912 |
| GO:0009451 | RNA modification: Biological Process | 1 | 20 | 0.01885091 | 0.049090912 |

**Table 3.** Enriched gene ontology terms for DMCs in protein coding genes within 1kb of an annotated repetitive element.

| GO ID | GO description | Subset count | Genome count | Raw p-value | Adj. p-value |
|---|---|---|---|---|---|
| GO:0005789 | endoplasmic reticulum membrane: Cellular Component | 2 | 19 | 0.000124423 | 0.003919194 |
| GO:0006467 | protein thiol-disulfide exchange: Biological Process | 1 | 3 | 7.97151E-05 | 0.003919194 |
| GO:0008250 | oligosaccharyltransferase complex: Cellular Component | 1 | 4 | 0.000158886 | 0.003919194 |
| GO:0004252 | serine-type endopeptidase activity: Molecular Function | 4 | 115 | 0.000340708 | 0.004170553 |
| GO:0004556 | alpha-amylase activity: Molecular Function | 1 | 5 | 0.000263907 | 0.004170553 |
| GO:0004579 | dolichyl-diphosphooligosaccharide-protein glycotransferase activity: Molecular Function | 1 | 6 | 0.000394512 | 0.004170553 |
| GO:0016671 | oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor: Molecular Function | 1 | 6 | 0.000394512 | 0.004170553 |
| GO:0004659 | prenyltransferase activity: Molecular Function | 1 | 11 | 0.001422087 | 0.013154302 |
| GO:0004519 | endonuclease activity: Molecular Function | 1 | 14 | 0.002328997 | 0.017234575 |
| GO:0016853 | isomerase activity: Molecular Function | 1 | 14 | 0.002328997 | 0.017234575 |
| GO:0005643 | nuclear pore: Cellular Component | 1 | 17 | 0.003445371 | 0.023177948 |
| GO:0015977 | carbon fixation: Biological Process | 1 | 18 | 0.003862895 | 0.023821184 |
| GO:0003725 | double-stranded RNA binding: Molecular Function | 1 | 21 | 0.005248288 | 0.025891553 |
| GO:0004629 | phospholipase C activity: Molecular Function | 1 | 20 | 0.004764592 | 0.025891553 |
| GO:0009536 | plastid: Cellular Component | 1 | 21 | 0.005248288 | 0.025891553 |
| GO:0003993 | acid phosphatase activity: Molecular Function | 1 | 26 | 0.007985868 | 0.035337434 |
| GO:0006629 | lipid metabolic process: Biological Process | 4 | 240 | 0.008435545 | 0.035337434 |
| GO:0008081 | phosphoric diester hydrolase activity: Molecular Function | 1 | 27 | 0.008595592 | 0.035337434 |
| GO:0010181 | FMN binding: Molecular Function | 1 | 28 | 0.009225524 | 0.035930987 |

**Table 4.** Frequency of DMCs in regions upstream of protein coding genes and within 1kb of a predicted repetitive element.

| mRNA id | Freq. DMC | Average % mC Sativa | Average %mC Serriola | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|
| Lsat_1_v5_gn_5_60241 | 35 | 69.92 | 22.87 | | |
| Lsat_1_v5_gn_4_127800 | 17 | 70.59 | 24.45 | | |
| Lsat_1_v5_gn_4_127821 | 17 | 70.59 | 24.45 | | |
| Lsat_1_v5_gn_8_116481 | 7 | 83.82 | 31.70 | GO:0006508: proteolysis: Biological Process | hypothetical protein LOC100242026 \| K01365 cathepsin L [EC:3.4.22.15] |
| Lsat_1_v5_gn_9_115440 | 7 | 58.03 | 7.02 | GO:0016651: oxidoreductase activity, acting on NADH or NADPH: Molecular Function \| GO:0048038: quinone binding: Molecular Function \| GO:0051287: NAD or NADH binding: Molecular Function | ndhH \| NADH dehydrogenase 49 kDa subunit \| K05579 NADH dehydrogenase I subunit 7 [EC:1.6.5.3] |
| Lsat_1_v5_gn_8_33421 | 7 | 97.10 | 19.11 | | nucleobase:cation symporter \| K03457 nucleobase:cation symporter-1, NCS1 family |
| Lsat_1_v5_gn_6_10001 | 7 | 73.87 | 2.94 | | |
| Lsat_1_v5_gn_5_137180 | 6 | 62.76 | 21.97 | GO:0005488: binding: Molecular Function \| GO:0005743: mitochondrial inner membrane: Cellular Component \| GO:0006810: transport: Biological Process | Hypothetical protein CBG03436 \| K05863 solute carrier family 25 (mitochondrial carrier \| adenine nucleotide translocator) |
| Lsat_1_v5_gn_7_83821 | 5 | 41.96 | 39.39 | GO:0004553: hydrolase activity, hydrolyzing O-glycosyl compounds: Molecular Function \| GO:0005618: cell wall: Cellular Component \| GO:0005975: carbohydrate metabolic process: Biological Process \| GO:0006073: cellular glucan metabolic process: Biological Process \| GO:0016762: xyloglucan:xyloglucosyl transferase activity: Molecular Function | Brassinosteroid-regulated protein BRU1 precursor, putative (EC:2.4.1.207) \| K08235 xyloglucan:xyloglucosyl transferase [EC:2.4.1.207] |
| Lsat_1_v5_gn_5_157381 | 4 | 39.08 | 21.86 | GO:0004553: hydrolase activity, hydrolyzing O-glycosyl compounds: Molecular Function | endo-1,4-beta-mannosidase (EC:3.-.-.-) \| K01567 [EC:3.-.-.-] |
| Lsat_1_v5_gn_2_3840 | 4 | 0.00 | 58.90 | | hypothetical protein LOC100243760 \| K02116 ATP synthase protein IScaffold=Lsat_1_v5_g_2_5839 |
| Lsat_1_v5_gn_9_51020 | 4 | 76.16 | 5.42 | GO:0009055: electron carrier activity: Molecular Function \| GO:0016020: | petB \| cytochrome b6 \| K02635 cytochrome b6 |

| mRNA id | Freq. DMC | Average % mC Sativa | Average %mC Serriola | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|
| | | | | membrane: Cellular Component \| GO:0016491: oxidoreductase activity: Molecular Function | |
| Lsat_1_v5_gn_8_159581 | 4 | 93.81 | 0.00 | GO:0006813: potassium ion transport: Biological Process \| GO:0015079: potassium ion transmembrane transporter activity: Molecular Function | potassium ion transporter family protein \| K03549 KUP system potassium uptake protein |
| Lsat_1_v5_gn_1_101 | 4 | 65.43 | 95.08 | GO:0006855: drug transmembrane transport: Biological Process \| GO:0015238: drug transmembrane transporter activity: Molecular Function \| GO:0015297: antiporter activity: Molecular Function \| GO:0016020: membrane: Cellular Component | |
| Lsat_1_v5_gn_3_1800 | 4 | 49.81 | 29.99 | | |
| Lsat_1_v5_gn_4_183200 | 4 | 73.71 | 26.02 | | |
| Lsat_1_v5_gn_7_100541 | 4 | 85.35 | 1.92 | | |
| Lsat_1_v5_gn_7_38301 | 3 | 33.67 | 2.98 | GO:0009522: photosystem I: Cellular Component \| GO:0009579: thylakoid: Cellular Component \| GO:0015979: photosynthesis: Biological Process | psaA \| photosystem I P700 chlorophyll a apoprotein A1 \| K02689 photosystem I P700 chlorophyll a apoprotein A1 |
| Lsat_1_v5_gn_2_89801 | 3 | 24.70 | 0.36 | GO:0000287: magnesium ion binding: Molecular Function \| GO:0009536: plastid: Cellular Component \| GO:0015977: carbon fixation: Biological Process | rbcL \| ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (EC:4.1.1.39) \| K01601 ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39] |
| Lsat_1_v5_gn_1_42401 | 3 | 86.31 | 5.12 | | |
| Lsat_1_v5_gn_4_10101 | 3 | 95.95 | 0.00 | | |
| Lsat_1_v5_gn_4_185040 | 3 | 64.93 | 30.30 | | |
| Lsat_1_v5_gn_4_186001 | 3 | 97.04 | 31.46 | | |
| Lsat_1_v5_gn_5_168120 | 3 | 36.13 | 7.31 | | |
| Lsat_1_v5_gn_5_60261 | 3 | 72.36 | 38.20 | | |
| Lsat_1_v5_gn_2_96380 | 2 | 85.10 | 54.17 | GO:0004672: protein kinase activity: Molecular Function \| GO:0004674: protein serine/threonine kinase activity: Molecular Function \| GO:0005524: | APK2A \| APK2A (PROTEIN KINASE 2A) \| ATP binding / kinase/ protein kinase/ protein serine/threonine kinase \| K00924 [EC:2.7.1.-] |

| mRNA id | Freq. DMC | Average % mC Sativa | Average %mC Serriola | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|
| | | | | ATP binding: Molecular Function | |
| Lsat_1_v5_gn_4_49540 | 2 | 92.61 | 5.56 | GO:0005215: transporter activity: Molecular Function \| GO:0006810: transport: Biological Process | Aquaporin PIP2.2, putative \| K09872 aquaporin PIP |
| Lsat_1_v5_gn_6_80221 | 2 | 75.46 | 0.00 | | ATLUP2 \| ATLUP2 \| beta-amyrin synthase/ lupeol synthase \| K01853 cycloartenol synthase [EC:5.4.99.8] |
| Lsat_1_v5_gn_3_73580 | 2 | 100.00 | 21.29 | GO:0004497: monooxygenase activity: Molecular Function \| GO:0005506: iron ion binding: Molecular Function \| GO:0009055: electron carrier activity: Molecular Function \| GO:0020037: heme binding: Molecular Function | CAld5H/F5H1, CYP84A10 \| coniferylaldehyde 5-hydroxylase \| K09755 ferulate-5-hydroxylase [EC:1.14.-.-] |
| Lsat_1_v5_gn_8_11480 | 2 | 90.42 | 14.71 | GO:0016872: intramolecular lyase activity: Molecular Function | Chalcone--flavonone isomerase, putative (EC:5.5.1.6) \| K01859 chalcone isomerase [EC:5.5.1.6] |
| Lsat_1_v5_gn_8_42261 | 2 | 96.03 | 30.00 | GO:0005622: intracellular: Cellular Component | CO \| CO (CONSTANS) \| transcription factor/ transcription regulator/ zinc ion binding \| K12135 zinc finger protein CONSTANS |
| Lsat_1_v5_gn_9_22560 | 2 | 78.71 | 30.68 | GO:0006508: proteolysis: Biological Process | hypothetical protein \| K01285 lysosomal Pro-X carboxypeptidase [EC:3.4.16.2] |
| Lsat_1_v5_gn_4_7040 | 2 | 60.20 | 48.81 | GO:0005488: binding: Molecular Function \| GO:0006886: intracellular protein transport: Biological Process \| GO:0016192: vesicle-mediated transport: Biological Process \| GO:0030117: membrane coat: Cellular Component \| GO:0030131: clathrin adaptor complex: Cellular Component | hypothetical protein \| K12391 AP-1 complex subunit gamma-1 |
| Lsat_1_v5_gn_4_175880 | 2 | 48.53 | 1.61 | GO:0003676: nucleic acid binding: Molecular Function | hypothetical protein LOC100247996 \| K13128 zinc finger CCHC domain-containing protein 8 |
| Lsat_1_v5_gn_8_36821 | 2 | 82.61 | 50.00 | GO:0003677: DNA binding: Molecular Function \| GO:0003899: DNA-directed RNA polymerase activity: Molecular Function | hypothetical protein LOC100263361 \| K10908 DNA-directed RNA polymerase, mitochondrial [EC:2.7.7.6] |
| Lsat_1_v5_gn_2_100901 | 2 | 100.00 | 0.00 | | leucine-rich repeat transmembrane protein kinase, putative (EC:1.3.1.74) \| K13420 LRR receptor-like |

| mRNA id | Freq. DMC | Average % mC Sativa | Average %mC Serriola | Gene Ontology | KEGG ID |
|---|---|---|---|---|---|
| | | | | | serine/threonine-protein kinase FLS2 [EC:2.7.11.1] |
| Lsat_1_v5_gn_1_111240 | 2 | 90.24 | 23.60 | | |
| Lsat_1_v5_gn_2_43920 | 2 | 92.26 | 9.03 | | |
| Lsat_1_v5_gn_2_62741 | 2 | 18.18 | 95.00 | GO:0003676: nucleic acid binding: Molecular Function \| GO:0004519: endonuclease activity: Molecular Function \| GO:0006308: DNA catabolic process: Biological Process | |
| Lsat_1_v5_gn_2_98441 | 2 | 34.44 | 1.63 | | |
| Lsat_1_v5_gn_3_112240 | 2 | 88.54 | 27.54 | | |
| Lsat_1_v5_gn_3_138461 | 2 | 97.20 | 1.03 | | |
| Lsat_1_v5_gn_4_176941 | 2 | 53.44 | 47.31 | | |
| Lsat_1_v5_gn_4_180340 | 2 | 93.73 | 32.14 | | |
| Lsat_1_v5_gn_4_82360 | 2 | 69.85 | 0.00 | | |
| Lsat_1_v5_gn_5_105761 | 2 | 43.40 | 2.60 | | |
| Lsat_1_v5_gn_5_109401 | 2 | 83.47 | 76.04 | | |
| Lsat_1_v5_gn_5_121221 | 2 | 38.69 | 4.12 | | |
| Lsat_1_v5_gn_5_128341 | 2 | 45.76 | 95.65 | | |
| Lsat_1_v5_gn_5_154760 | 2 | 87.33 | 0.00 | | |

**Table 5.** Enriched gene ontology terms for DMCs upstream of protein coding genes and within 1 kb of an annotated repetitive element.

| GO ID | GO description | Subset count | Genome count | Raw p-value | Adj. p-value |
|---|---|---|---|---|---|
| GO:0006073 | cellular glucan metabolic process: Biological Process | 1 | 1 | 0 | 0 |
| GO:0016762 | xyloglucan:xyloglucosyl transferase activity: Molecular Function | 1 | 1 | 0 | 0 |
| GO:0030131 | clathrin adaptor complex: Cellular Component | 1 | 1 | 0 | 0 |
| GO:0009522 | photosystem I: Cellular Component | 2 | 18 | 3.40369E-05 | 0.000470334 |
| GO:0009579 | thylakoid: Cellular Component | 2 | 19 | 4.03144E-05 | 0.000470334 |
| GO:0051287 | NAD or NADH binding: Molecular Function | 1 | 3 | 3.72573E-05 | 0.000470334 |
| GO:0006308 | DNA catabolic process: Biological Process | 1 | 4 | 7.43415E-05 | 0.000743415 |
| GO:0005742 | mitochondrial outer membrane translocase complex: Cellular Component | 1 | 6 | 0.000184992 | 0.001618681 |
| GO:0016872 | intramolecular lyase activity: Molecular Function | 1 | 8 | 0.000343718 | 0.002673365 |
| GO:0009521 | photosystem: Cellular Component | 1 | 10 | 0.000549846 | 0.003207436 |
| GO:0009767 | photosynthetic electron transport chain: | 1 | 10 | 0.000549846 | 0.003207436 |

| GO ID | GO description | Subset count | Genome count | Raw p-value | Adj. p-value |
|---|---|---|---|---|---|
| | Biological Process | | | | |
| GO:0016168 | chlorophyll binding: Molecular Function | 1 | 10 | 0.000549846 | 0.003207436 |
| GO:0016020 | membrane: Cellular Component | 7 | 539 | 0.000690914 | 0.00345457 |
| GO:0030117 | membrane coat: Cellular Component | 1 | 11 | 0.000670477 | 0.00345457 |
| GO:0015979 | photosynthesis: Biological Process | 2 | 51 | 0.00079771 | 0.003722646 |
| GO:0004519 | endonuclease activity: Molecular Function | 1 | 14 | 0.001101644 | 0.00453618 |
| GO:0016853 | isomerase activity: Molecular Function | 1 | 14 | 0.001101644 | 0.00453618 |
| GO:0031461 | cullin-RING ubiquitin ligase complex: Cellular Component | 1 | 15 | 0.001268185 | 0.004931829 |
| GO:0004190 | aspartic-type endopeptidase activity: Molecular Function | 2 | 71 | 0.002079611 | 0.006616145 |
| GO:0005743 | mitochondrial inner membrane: Cellular Component | 2 | 72 | 0.002164466 | 0.006616145 |
| GO:0015079 | potassium ion transmembrane transporter activity: Molecular Function | 1 | 20 | 0.002268393 | 0.006616145 |
| GO:0015977 | carbon fixation: Biological Process | 1 | 18 | 0.001835129 | 0.006616145 |
| GO:0016651 | oxidoreductase activity, acting on NADH or NADPH: Molecular Function | 1 | 20 | 0.002268393 | 0.006616145 |
| GO:0048038 | quinone binding: Molecular Function | 1 | 20 | 0.002268393 | 0.006616145 |
| GO:0009536 | plastid: Cellular Component | 1 | 21 | 0.002501376 | 0.006734475 |
| GO:0016192 | vesicle-mediated transport: Biological Process | 2 | 75 | 0.002431838 | 0.006734475 |
| GO:0004252 | serine-type endopeptidase activity: Molecular Function | 2 | 115 | 0.008027475 | 0.020068686 |
| GO:0005975 | carbohydrate metabolic process: Biological Process | 2 | 114 | 0.007838015 | 0.020068686 |
| GO:0004970 | ionotropic glutamate receptor activity: Molecular Function | 1 | 41 | 0.009326821 | 0.021762582 |
| GO:0005234 | extracellular-glutamate-gated ion channel activity: Molecular Function | 1 | 41 | 0.009326821 | 0.021762582 |
| GO:0004221 | ubiquitin thiolesterase activity: Molecular Function | 1 | 46 | 0.011637708 | 0.025457487 |
| GO:0042802 | identical protein binding: Molecular Function | 1 | 46 | 0.011637708 | 0.025457487 |
| GO:0006511 | ubiquitin-dependent protein catabolic process: Biological Process | 1 | 48 | 0.012625282 | 0.025535109 |
| GO:0006855 | drug transmembrane transport: Biological Process | 1 | 49 | 0.013132342 | 0.025535109 |
| GO:0015238 | drug transmembrane transporter activity: Molecular Function | 1 | 49 | 0.013132342 | 0.025535109 |
| GO:0015297 | antiporter activity: Molecular Function | 1 | 49 | 0.013132342 | 0.025535109 |
| GO:0004650 | polygalacturonase activity: Molecular Function | 1 | 51 | 0.014172668 | 0.026107546 |
| GO:0016887 | ATPase activity: Molecular Function | 2 | 141 | 0.013916917 | 0.026107546 |
| GO:0006810 | transport: Biological Process | 4 | 413 | 0.016017268 | 0.028030219 |
| GO:0006813 | potassium ion transport: Biological Process | 1 | 54 | 0.015797668 | 0.028030219 |
| GO:0004553 | hydrolase activity, hydrolyzing O-glycosyl compounds: Molecular Function | 3 | 275 | 0.016766635 | 0.028625963 |
| GO:0006508 | proteolysis: Biological Process | 3 | 300 | 0.022257123 | 0.037095205 |
| GO:0003899 | DNA-directed RNA polymerase activity: Molecular Function | 1 | 66 | 0.023038469 | 0.037504484 |

| GO ID | GO description | Subset count | Genome count | Raw p-value | Adj. p-value |
|---|---|---|---|---|---|
| GO:0005618 | cell wall: Cellular Component | 1 | 73 | 0.027778767 | 0.044193493 |

**Table 6.** Gene ontology analysis of genes containing highly conserved methylation among replicates and between genotypes.

| GO ID | GO description | Subset count | Genome count | Raw p-value | Adj. p-value |
|---|---|---|---|---|---|
| GO:0006123 | mitochondrial electron transport, cytochrome c to oxygen: Biological Process | 2 | 2 | 0 | 0 |
| GO:0008289 | lipid binding: Molecular Function | 1 | 1 | 0 | 0 |
| GO:0015986 | ATP synthesis coupled proton transport: Biological Process | 3 | 24 | 3.20849E-07 | 5.88223E-06 |
| GO:0015078 | hydrogen ion transmembrane transporter activity: Molecular Function | 3 | 32 | 1.07E-06 | 1.47125E-05 |
| GO:0033177 | proton-transporting two-sector ATPase complex, proton-transporting domain: Cellular Component | 2 | 13 | 3.79235E-06 | 4.17158E-05 |
| GO:0004129 | cytochrome-c oxidase activity: Molecular Function | 2 | 14 | 4.81823E-06 | 4.41671E-05 |
| GO:0051287 | NAD or NADH binding: Molecular Function | 1 | 3 | 1.71898E-05 | 0.000135063 |
| GO:0009507 | chloroplast: Cellular Component | 1 | 6 | 8.55458E-05 | 0.00052278 |
| GO:0042773 | ATP synthesis coupled electron transport: Biological Process | 1 | 6 | 8.55458E-05 | 0.00052278 |
| GO:0003777 | microtubule motor activity: Molecular Function | 2 | 67 | 0.000578356 | 0.002164938 |
| GO:0008137 | NADH dehydrogenase (ubiquinone) activity: Molecular Function | 1 | 13 | 0.000439986 | 0.002164938 |
| GO:0009772 | photosynthetic electron transport in photosystem II: Biological Process | 1 | 14 | 0.000512514 | 0.002164938 |
| GO:0015991 | ATP hydrolysis coupled proton transport: Biological Process | 2 | 60 | 0.000418179 | 0.002164938 |
| GO:0019684 | photosynthesis, light reaction: Biological Process | 1 | 15 | 0.000590438 | 0.002164938 |
| GO:0030077 | plasma membrane light-harvesting complex: Cellular Component | 1 | 14 | 0.000512514 | 0.002164938 |
| GO:0016307 | phosphatidylinositol phosphate kinase activity: Molecular Function | 1 | 16 | 0.000673731 | 0.002315949 |
| GO:0016651 | oxidoreductase activity, acting on NADH or NADPH: Molecular Function | 1 | 20 | 0.001060086 | 0.003239151 |
| GO:0048038 | quinone binding: Molecular Function | 1 | 20 | 0.001060086 | 0.003239151 |
| GO:0008375 | acetylglucosaminyltransferase activity: Molecular Function | 1 | 34 | 0.003062342 | 0.008864674 |
| GO:0005576 | extracellular region: Cellular Component | 1 | 35 | 0.003242876 | 0.008917909 |
| GO:0006869 | lipid transport: Biological Process | 1 | 46 | 0.005545107 | 0.013862768 |

| GO ID | GO description | Subset count | Genome count | Raw p-value | Adj. p-value |
|---|---|---|---|---|---|
| GO:0042802 | identical protein binding: Molecular Function | 1 | 46 | 0.005545107 | 0.013862768 |
| GO:0015979 | photosynthesis: Biological Process | 1 | 51 | 0.006777992 | 0.016208243 |
| GO:0003735 | structural constituent of ribosome: Molecular Function | 3 | 331 | 0.008488664 | 0.018675061 |
| GO:0005840 | ribosome: Cellular Component | 3 | 331 | 0.008488664 | 0.018675061 |
| GO:0004190 | aspartic-type endopeptidase activity: Molecular Function | 1 | 71 | 0.01280657 | 0.027090822 |

**Table 7.** Enriched gene ontology terms for protein coding genes containing DVC.

| GO ID | GO description | Subset count | Genome count | Raw p-value | Adj. p-value |
|---|---|---|---|---|---|
| GO:0006123 | mitochondrial electron transport, cytochrome-c to oxygen: Biological Process | 2 | 2 | 0 | 0 |
| GO:0009060 | aerobic respiration: Biological Process | 1 | 1 | 0 | 0 |
| GO:0004129 | cytochrome-c oxidase activity: Molecular Function | 2 | 14 | 6.54545E-07 | 7.41817E-06 |
| GO:0003964 | RNA-directed DNA polymerase activity: Molecular Function | 1 | 6 | 2.28035E-05 | 0.000155064 |
| GO:0042773 | ATP synthesis coupled electron transport: Biological Process | 1 | 6 | 2.28035E-05 | 0.000155064 |
| GO:0006461 | protein complex assembly: Biological Process | 1 | 9 | 5.45983E-05 | 0.00020626 |
| GO:0008535 | respiratory chain complex IV assembly: Biological Process | 1 | 8 | 4.24991E-05 | 0.00020626 |
| GO:0015232 | heme transporter activity: Molecular Function | 1 | 9 | 5.45983E-05 | 0.00020626 |
| GO:0015886 | heme transport: Biological Process | 1 | 9 | 5.45983E-05 | 0.00020626 |
| GO:0008137 | NADH dehydrogenase (ubiquinone) activity: Molecular Function | 1 | 13 | 0.000117922 | 0.000400933 |
| GO:0015986 | ATP synthesis coupled proton transport: Biological Process | 1 | 24 | 0.000413637 | 0.001278515 |
| GO:0015078 | hydrogen ion transmembrane transporter activity: Molecular Function | 1 | 32 | 0.000738651 | 0.002092844 |
| GO:0015979 | photosynthesis: Biological Process | 1 | 51 | 0.001870407 | 0.004891834 |
| GO:0003777 | microtubule motor activity: Molecular Function | 1 | 67 | 0.003202742 | 0.007259549 |
| GO:0003899 | DNA-directed RNA polymerase activity: Molecular Function | 1 | 66 | 0.003109594 | 0.007259549 |
| GO:0004190 | aspartic-type endopeptidase activity: Molecular Function | 1 | 71 | 0.00358829 | 0.007625117 |
| GO:0003735 | structural constituent of ribosome: Molecular Function | 2 | 331 | 0.00818573 | 0.015461935 |
| GO:0005840 | ribosome: Cellular Component | 2 | 331 | 0.00818573 | 0.015461935 |
| GO:0006412 | translation: Biological Process | 1 | 119 | 0.009761785 | 0.017468457 |
| GO:0003723 | RNA binding: Molecular Function | 1 | 197 | 0.025249811 | 0.04129632 |
| GO:0005506 | iron ion binding: Molecular Function | 2 | 508 | 0.025506551 | 0.04129632 |
| GO:0009055 | electron carrier activity: Molecular Function | 2 | 534 | 0.028984986 | 0.044794979 |

**Table 8.** Genes containing at least one cytosine which is covered by at least 10 reads and fully methylated in each biological replicate of both *L. sativa* and *L. serriola*.

| Chromosome | Start position | End position | Context | Gene ID | GO ID | KEGG ID |
|---|---|---|---|---|---|---|
| Lsat_1_v6_lg_4 | 350455078 | 350458757 | CG | Lsat_1_v5_gn_4_152881 | GO:0004672: protein kinase activity: Molecular Function \| GO:0004674: protein serine/threonine kinase activity: Molecular Function \| GO:0005515: protein binding: Molecular Function \| GO:0005524: ATP binding: Molecular Function | FLS2 \| FLS2 (FLAGELLIN-SENSITIVE 2) \| ATP binding / kinase/ protein binding / protein serine/threonine kinase/ transmembrane receptor protein serine/threonine kinase \| K13420 LRR receptor-like serine/threonine-protein kinase FLS2 [EC:2.7.11.1] |
| Lsat_1_v6_lg_4 | 350455078 | 350458757 | CHG | Lsat_1_v5_gn_4_152881 | GO:0004672: protein kinase activity: Molecular Function \| GO:0004674: protein serine/threonine kinase activity: Molecular Function \| GO:0005515: protein binding: Molecular Function \| GO:0005524: ATP binding: Molecular Function | FLS2 \| FLS2 (FLAGELLIN-SENSITIVE 2) \| ATP binding / kinase/ protein binding / protein serine/threonine kinase/ transmembrane receptor protein serine/threonine kinase \| K13420 LRR receptor-like serine/threonine-protein kinase FLS2 [EC:2.7.11.1] |
| Lsat_1_v6_lg_4 | 394333444 | 394338489 | CHH | Lsat_1_v5_gn_4_167780 | | |
| Lsat_1_v6_lg_4 | 394333444 | 394338489 | CHH | Lsat_1_v5_gn_4_167780 | | |
| Lsat_1_v6_lg_4 | 394333444 | 394338489 | CG | Lsat_1_v5_gn_4_167780 | | |
| Lsat_1_v6_lg_4 | 394333444 | 394338489 | CHG | Lsat_1_v5_gn_4_167780 | | |
| Lsat_1_v6_lg_4 | 394333444 | 394338489 | CHG | Lsat_1_v5_gn_4_167780 | | |
| Lsat_1_v6_lg_4 | 394333444 | 394338489 | CHH | Lsat_1_v5_gn_4_167780 | | |
| Lsat_1_v6_lg_4 | 394333444 | 394338489 | CHG | Lsat_1_v5_gn_4_167780 | | |
| Lsat_1_v6_lg_4 | 394333444 | 394338489 | CHG | Lsat_1_v5_gn_4_167780 | | |
| Lsat_1_v6_lg_4 | 394333444 | 394338489 | CHG | Lsat_1_v5_gn_4_167780 | | |
| Lsat_1_v6_lg_4 | 432409453 | 432409740 | CG | Lsat_1_v5_gn_4_183940 | | |
| Lsat_1_v6_lg_7 | 191848981 | 191850978 | CHG | Lsat_1_v5_gn_7_96181 | | |
| Lsat_1_v6_lg_7 | 191848981 | 191850978 | CHH | Lsat_1_v5_gn_7_96181 | | |
| Lsat_1_v6_lg_7 | 191848981 | 191850978 | CHG | Lsat_1_v5_gn_7_96181 | | |
| Lsat_1_v6_lg_7 | 191848981 | 191850978 | CHH | Lsat_1_v5_gn_7_96181 | | |
| Lsat_1_v6_lg_9 | 50682250 | 50685381 | CHG | Lsat_1_v5_gn_9_38740 | | |
| Lsat_1_v6_lg_9 | 50682250 | 50685381 | CG | Lsat_1_v5_gn_9_38740 | | |

| Chromosome | Start position | End position | Context | Gene ID | GO ID | KEGG ID |
|---|---|---|---|---|---|---|
| Lsat_1_v6_lg_9 | 50682250 | 50685381 | CG | Lsat_1_v5_gn_9_38740 | | |
| Lsat_1_v6_lg_9 | 50682250 | 50685381 | CG | Lsat_1_v5_gn_9_38740 | | |
| Lsat_1_v6_lg_9 | 50682250 | 50685381 | CG | Lsat_1_v5_gn_9_38740 | | |
| Lsat_1_v6_lg_9 | 50682250 | 50685381 | CG | Lsat_1_v5_gn_9_38740 | | |
| Lsat_1_v6_lg_9 | 50682250 | 50685381 | CG | Lsat_1_v5_gn_9_38740 | | |
| Lsat_1_v6_lg_9 | 50682250 | 50685381 | CG | Lsat_1_v5_gn_9_38740 | | |
| Lsat_1_v6_lg_9 | 50682250 | 50685381 | CG | Lsat_1_v5_gn_9_38740 | | |
| Lsat_1_v6_lg_9 | 88567292 | 88574707 | CG | Lsat_1_v5_gn_9_60620 | | |
| Lsat_1_v6_lg_9 | 88567292 | 88574707 | CG | Lsat_1_v5_gn_9_60620 | | |
| Lsat_1_v6_lg_9 | 88567292 | 88574707 | CG | Lsat_1_v5_gn_9_60620 | | |
| Lsat_1_v6_lg_9 | 88567292 | 88574707 | CG | Lsat_1_v5_gn_9_60620 | | |
| Lsat_1_v6_lg_9 | 88567292 | 88574707 | CHG | Lsat_1_v5_gn_9_60620 | | |
| Lsat_1_v6_lg_9 | 88567292 | 88574707 | CG | Lsat_1_v5_gn_9_60620 | | |
| Lsat_1_v6_lg_9 | 88567292 | 88574707 | CHG | Lsat_1_v5_gn_9_60620 | | |
| Lsat_1_v6_lg_9 | 88567292 | 88574707 | CG | Lsat_1_v5_gn_9_60620 | | |
| Lsat_1_v6_lg_9 | 88567292 | 88574707 | CG | Lsat_1_v5_gn_9_60620 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CHH | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CHH | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CHH | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CG | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CG | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CG | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CHH | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CHH | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CHH | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CHH | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CHH | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CG | Lsat_1_v5_gn_9_80520 | | |
| Lsat_1_v6_lg_9 | 143112314 | 143121896 | CHH | Lsat_1_v5_gn_9_80520 | | |

# CHAPTER 3

# MODIFIED REDUCED REPRESENTATION BISULFITE SEQUENCING FOR PLANT GENOMES

## Introduction

Reduced representation bisulfite sequencing (RRBS) allows a cost-effective whole genome survey suitable for assessing methylation in large genomes, and for assessing methylation variation across populations of individuals, tissues and treatments. The majority of RRBS studies, including the one example of RRBS of a plant genome [135], utilize the restriction endonuclease MspI. MspI is the most widely used enzyme in RRBS as its cleavage is insensitive to methylation in the predominant context for mammalian methylation, CG. However, MspI cleavage is blocked by methylation in the outer C of its recognition sequence (mCCGG). Unlike in mammalian species, the CHG (H = C, A, or T) context can account for a significant percentage of genome wide methylation in plants [23,25–28,30,54,107,136]. Methylation of proximal cytosines have been shown to be correlated [25,136], introducing the possibility that use of MspI in plants would bias calls of differential methylation.

**Methods**

To determine if this was the case we performed *in silico* digests of *Arabidopsis* genomes with known methylation states using MspI, a hypothetical MspI unaffected by methylation in its recognition site, and BssSI (CˆACGAG) and BsoBI (CˆYCGRG). BssSI and BsoBI are insensitive to methylation in their recognition sequences and generate 5' overhangs that are filled in during the end-repair step of library preparation providing an internal control for bisulfite non-conversion. We based the *in silico* digestions of *Arabidopsis thaliana* on an existing data set to most realistically model the variability in read coverage and methylation distribution through the genome. The data set generated by Dowen et al. (2012) included four libraries: two biological replicates in control condition and two biological replicated treated with salicylic acid [40]. Reads were trimmed to remove adapters using Trimmomatic (v 0.32) [137], aligned to the *Arabidopsis thaliana* genome (TAIR 10) using Bismark (v 0.13.0) [98] and Bowtie2 (v 2-2.1.0) [100]. The number of C's and T's in the reads covering each reference cytosine were used as the reference methylation state for generating the *in silico* RRBS libraries.

*In silico* libraries were generated by identifying MspI, BssSI, and BsoBI recognition sequences located 200 to 700 bp apart in the reference genome; these positions flank the potential sequence fragments in the *in silico* RRBS library. For libraries representing MspI's actual cleavage, the total number of reads for each fragment was equal to the minimum number of methylated reads covering the outer C's of the fragment's flanking restriction sites. For libraries representing an unbiased MspI digest and a double digest with BssSI and BsoBI the total number of reads for each fragment

was equal to the minimum number of total reads covering the outer C's of the fragment's flanking restriction sites. The proportion of methylation for each C within the fragment was taken to be the mode of 1,000 samplings from a binomial distribution with p = proportion of all reads that were methylated in the reference methylation library for that position and n = the total number of reads for the fragment. To detect differential methylation we removed from consideration positions which were covered by fewer than 4 reads in the whole genome data set, and positions where the variance in proportion of methylation between treatments was less than the 25 percentile. For each library, processing the CG, CHG and CHH contexts independently, we used the R function *prop.test* from the *stats* package to test each covered position for significant differences in the percent methylation in the between control and treatment conditions. Multiple test correction was performed using the method of Benjamini and Hochberg [138]. Positions with adjusted p-values less than 0.05 were considered to be differentially methylated cytosines (DMCs).

**Results and discussion**

*In silico analysis*

*In silico* libraries covered 658,776, 1,919,848 and 870,462 in methylation-sensitive MspI, methylation-insensitive MspI, and BssSI/BsoBI libraries respectively, representing 2.5%, 7.3% and 3.3% of cytosines covered by at least 4 reads in all replicates in the WGBS reference (Figure 1). We identified DMC between the salicylic acid treatment and control conditions in the methylation-sensitive MspI, methylation-

insensitive MspI, and BssSI/BsoBI *in silico* RRBS libraries. The BssSI/BsoBI digest covered slightly more genomic cytosines than the methylation-sensitive MspI digest, and the BssSI/BsoBI digest detected a much larger percentage of the DMC detected from the same genomics positions using the WGBS libraries (Figure 3). The *in silico* libraries representing the actual methylation-sensitive behavior of MspI detects 38-44% - 13% of the DMC detected using the same positions from the WGBS libraries, whereas the DMC detected in BssSI/BsoBI libraries represent 63-75% of DMC detected from the same positions in the WGBS libraries (Figure 3A). We repeated the analysis with a minimum read depth 10 reads per position, again more DMC were detected in BsoBI/BssSI RRBS libraries than actual MspI libraries (Figure 3B). However, the percentage of WGBS DMC detected in RRBS libraries varied with read depth and was lower at the higher read depth. DMC which were detected only in the RRBS libraries were found at positions with significant differences in the read coverage between DMC and RRBS libraries. There was no apparent sequence context bias between DMCs detected in RRBS-only, DMCs detected in both libraries or DMCs detected in only WGBS.

*Reduced representation bisulfite sequencing of L. serriola*

Additionally, we performed RRBS in the common weed *Lactuca serriola* using a double digest of BssSI and BsoBI. We generated libraries for each of three biological replicates of *L. serriola* grown without fertilizer and parental plants grown in control conditions (CN1, CN2, and CN3), and three biological replicates grown for two generations in without fertilizer (NN1, NN2, and NN3). DNA extractions and library construction methods are fully described in [139], important modifications to standard

protocols are noted here: 1.4 $\mu$g of purified DNA was digested for 10 hours with 30 units of BssSI and 30 units of BsoBI (New England Biolabs, Ipswich, MA), enzymes were heat inactivated and reactions cleaned-up, prior to library preparation using NEBNext® UltraTM DNA Library Prep Kit for Illumina® and NEBNext® Multiplex Oligos for Illumina® (Methylated Adaptors) according to manufacturer's instructions (New England Bio- labs, Ipswich, MA). Ligation products were purified and libraries were size selected with a 1.5% Blue Pippin agarose gel cassette for fragment sizes between 250-600 bp (Sage Science, Beverly, MA). Each library was PCR amplified for 13 cycles in a single 50 $\mu$l reaction containing 20 $\mu$l of bisulfite treated sample. Paired-end sequencing (2x100) was performed on a HiSeq 2000 at the University of Massachusetts Boston Center for Personalized Cancer Therapy Genomics Core. Sequence QC and alignment methods are fully described in [139].

There were 1,325,486 cytosines that were covered by at least ten reads in all replicates, representing 1% of all cytosines in the *L. serriola* genome, and 62% of cytosines covered by at least 10 reads in WGBS of *L. serriola* [136,139]. Mean methylation levels of RRBS libraries were consistent between biological replicates in both conditions (Figure 4) and with levels reported in WGBS of *L. serriola* [136]. The majority of covered cytosines are found in the CHH context 62%, with 22% and 16% of cytosines found in the CG and CHG contexts respectively. The majority of methylated cytosines were located in repetitive regions (84%), 4% of cytosines were found in protein coding genes and 10% in unannotated regions (Figure 6). Though mean methylation levels did not appreciably differ between the two treatment conditions, the samples did

contain distinguishing differences in proportion methylation across covered cytosine positions as shown by hierarchical clustering (Figure 5). RRBS provides a significant cost-savings over WGBS while providing reproducible information on the methylation status of a significant subset the plant genome. However, the choice of enzymes is critical to the success of this technique. Here we introduce a method of selecting restriction endonucleases suitable for use in RRBS of plant genomes. We show *in silico* data predicting improved performance of RRBS using BssSI and BsoBI relative to the traditional enzyme of choice, MspI, and performed RRBS of *L. serriola* using BssSI and BsoBI covering with at least ten reads and 62% of cytosines covered by at least 10 reads in WGBS of *L. serriola* with reproducible methylation values between biological replicates.

## Figures

**Figure 1.** Number of cytosines covered by *in silico* MspI and BssSI/BsoBI libraries. The number of cytosines in *Arabidopsis thaliana* covered in 200-700 bp fragments were calculated for *in silico* digests given MspI's actual cleavage, theoretical digest with an MspI not blocked by methylation in the $C^{m}CG$ context, and double digest of BssSI and BsoBI.

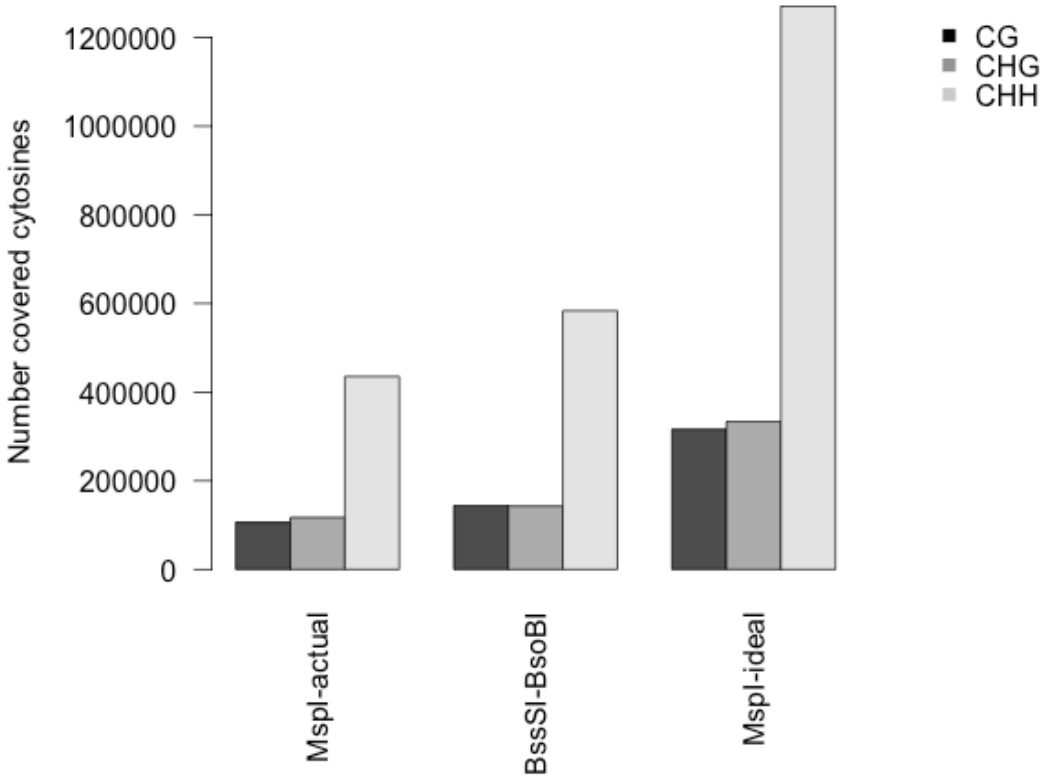


**Figure 2.** Distribution of cytosines in gene bodies, repetitive elements or other genomic regions for *in silico* digests with MspI, ideal MspI , BssSI and BsoBI.

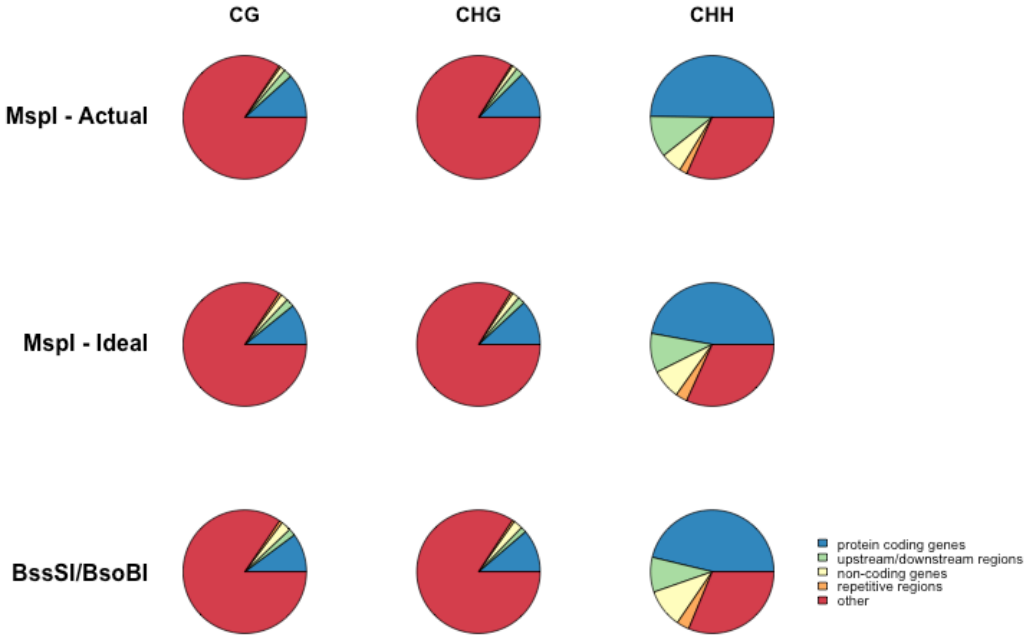
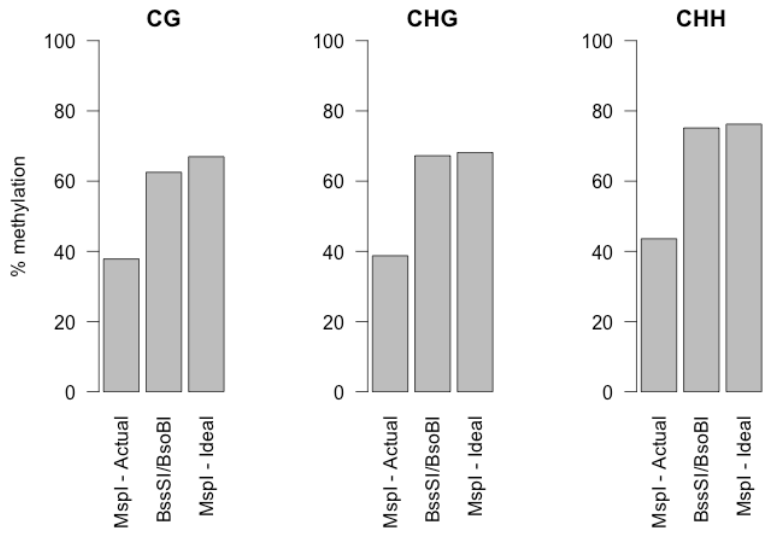


**Figure 3.** Percentage of significant DMCs by sequence context in *in silico* libraries. The percentage of significantly differentially methylated cytosines by sequence context was calculated for *in silico* MspI methylation sensitive libraries, MspI methylation insensitive libraries, and BssSI/BsoBI libraries relative to those detected in WGBS libraries considering positions found in both the RRBS and WGBS libraries.

**A.**




**B.**

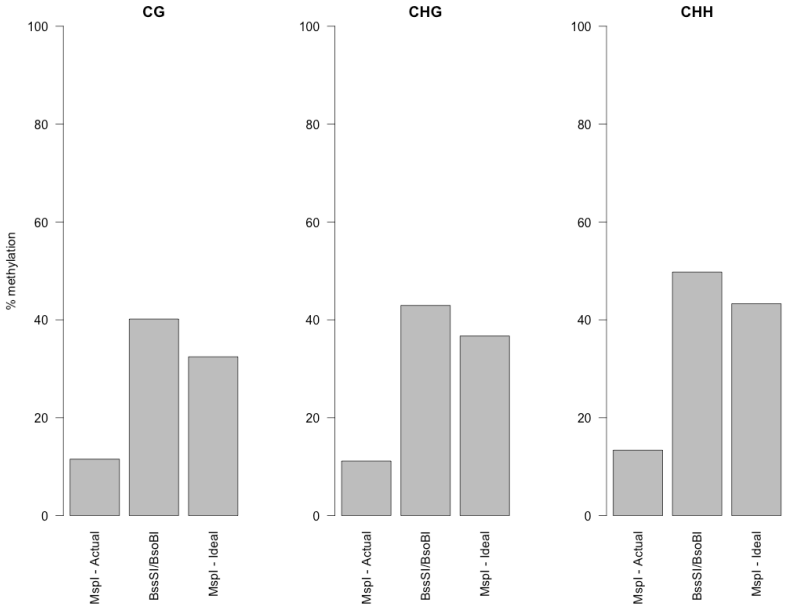


**Figure 4.** Median genome-wide levels of methylation by sequence context for RRBS of *L. serriola*. Samples *L. serriola* plants were grown without fertilizer and parental plants grown in control conditions (CN1, CN2, and CN3) or without fertilizer (NN1, NN2, and NN3).

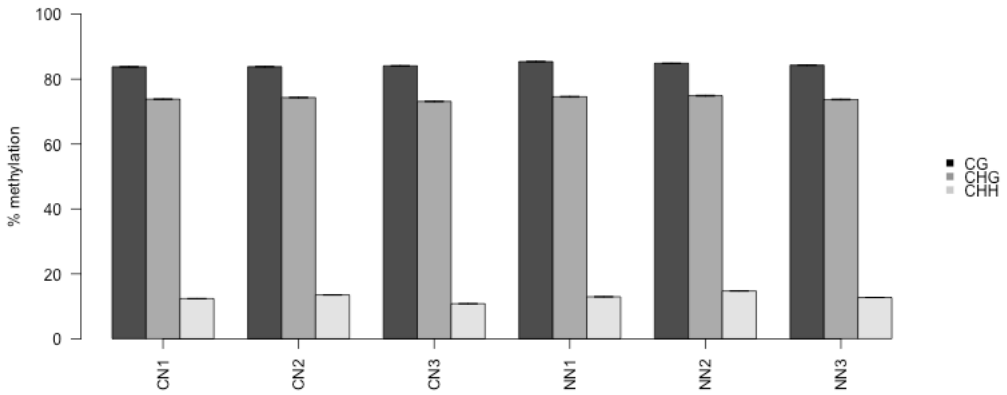


**Figure 5.** Hierarchical clustering of methylation in the CG (A), CHG (B), and CHH (C) contexts of *L. serriola* grown in different treatment conditions. In red are approximately unbiased probabilities generated by pvclust, in green are boot strap probabilities. CN1, CN2, and CN3 represent biological replicates grown in without fertilizer and parental plants grown in control conditions; N1, N2, and N3 represent biological replicates grown in no-fertilizer conditions.

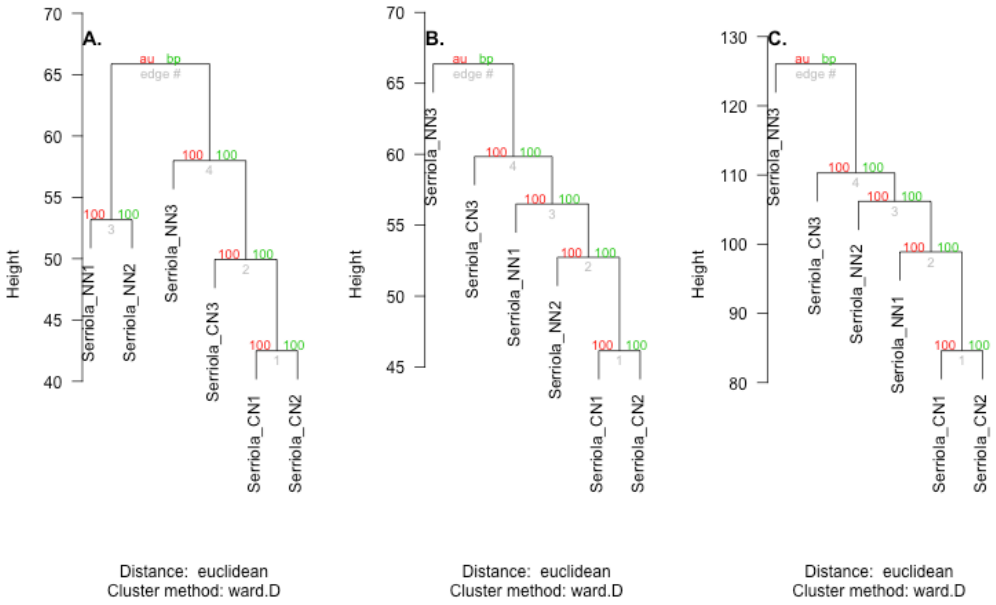
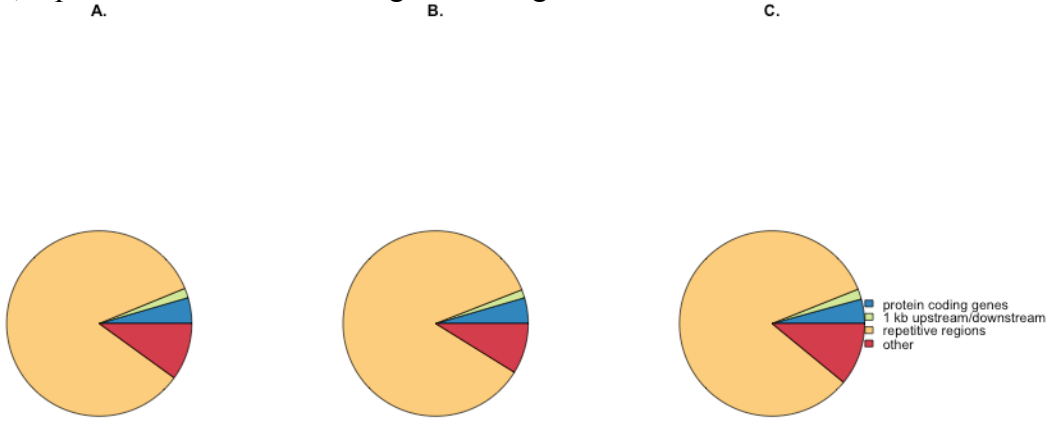


**Figure 6.** Distribution of cytosines in *L. serriola* by genomic region. The proportion of cytosine positions covered with at least 10 reads in all replicates and conditions was determined for the CG (A), CHG (B), and CHH (C) sequence contexts in *L. serriola* gene bodies, repetitive elements or other genomic regions.



# CHAPTER 4

# REDUCED REPRESENTATION BISULFITE SEQUENCING OF *L. SERRIOLA* AND *L. SATIVA* SALINAS WITH DIFFERING FAMILY HISTORIES OF NUTRIENT DEPRIVATION

## Introduction

Transmission of specific methylation in response to a stress has led some to suggest that DNA methylation is a mechanism for Lamarkian inheritance of acquired characteristics [140], while others suggest that acquired methylation serves to maintain phenotypic stochasticity in genetically homogeneous populations [74]. Both the Lamarkian and the stochastic variation models for the evolutionary effect of methylation could have profound consequences for the evolution of plants. In a Lamarkian sense, transmission of acquired methylation may pre-adapt offspring to the environment and result in a competitive advantage for individuals. Inheritance of stress associated DNA methylation could be directed towards the stress that is encountered. There are several examples in plants of transgenerational priming, improved fitness of offspring of stressed parents, including examples in radishes [141], *Arabidopsis* [42,46,142], monkey flowers [143], tomato [42], and tobacco [62]. In some cases these transgenerational benefits can

be reduced or eliminated in plants deficient in RNA directed DNA methylation or treated with a methylation inhibitor [42,46]. Alternately stochastic variation of methylation could explain the competitive success of invasive plants in highly disturbed environments. Feinberg & Irizarry (2010) modeled evolutionary consequences under different selection conditions and found that within a fixed environment the genotype with the greatest expected value for the desirable trait and the lowest stochastic variation was favored but in a variable environment the highly variable genotype was favored [74]. Indeed, Latzel et al. (2013) found that epigenetic diversity in *Arabidopsis* was associated with increased plant productivity (biomass) in environments challenged with plant competitors and pathogens [75].

Plants methylomes show significant effects of environmental stimuli. Natural populations of clonal plants have shown significant differences in methylation in different environments. For example, Gao et al. (2010) found differentially methylated loci in individual alligator weed clones grown in aquatic and terrestrial "common gardens" habitats regardless of the particular geography and habitat from which the plant was originally collected, suggesting that DNA methylation plays a role in adapting individuals to diverse habitats [144]. Results of controlled abiotic stress treatments have been very specific to the plant variety and experimental design. The duration of the stress and the time between when the plant experienced the stress and when the tissue was sampled may be important, but largely unconsidered, variables when relating different stress methylation studies, as methylation changes can be induced within a few hours of stress treatment [45,47]

and a large proportion of induced changes may revert with time [48]. For example, salt treatment of rice varieties showed global hypomethlyation in both sensitive and resistant varieties [44,45], though the degree and speed at which the changes accumulated differed by variety [45]. In contrast, salt stressed *A. thaliana* showed global hypermethylation, and local hypomethylation of abiotic stress response genes, and progeny of salt stressed plants showed increased germination and root length when grown in salt media [46].

Here we use RRBS to compare the acquisition and inheritance of methylcytosine between two *Lactuca* sp. with differing abilities to adapt to disturbed environments. Modern production of commercial lettuce requires moderate rates of nitrogen and phosphorous application. The domesticated variety *L. sativa* cv. Salinas was developed by the USDA in the 1980s and is the one of the mostly widely used elite cultivars in the breeding of modern crisphead lettuce varieties [93]. *Lactuca serriola* is a hardy weed commonly found beside highways and in other human disturbed environments. *L. serriola* is found on all continents with the exception of Antarctica, and is a common weed found throughout the lower 48 states [83]. The accession of *L. serriola* used in the present work (UC96US23) was originally derived from a plant growing in the parking lot of an abandoned gas station in Davis, CA.

## Methods

*Plant Growth, sample collection and DNA extraction*

*Lactuca sativa* cv. Salinas and *Lactuca serriola* (UC96US23) seeds were obtained from the Richard Michelmore at the University of California Davis and the Compositae Genome Project (http://compgenomics.ucdavis.edu). To reduce variation due to the maternal effect of different growing conditions for the different sources of seeds used in this study, a progenitor generation was planted, grown and self-pollinated prior to the start of this experiment. Phenotypic measures of the next "parental" generation (S0) were collected. Plants were bagged to ensure self-fertilization, and seeds collected. Each biological replicate in the offspring generation (S1) was derived from a different individual in the parental generation. Seeds from this S1 generation were sterilized according to the following procedure: 1 mL 20% bleach solution and one drop Tween 20 were added to 25 seeds in a 2 mL microcentrifuge tube and gently agitated for 5 minutes. After a quick spin, detergent solution was decanted and 1 mL autoclaved, deionized water added and tubes gently agitated for 5 minutes. This process was repeated for a total of 10 rinses. The seeds were refrigerated overnight at 4°C. Seeds were planted in commercial potting soil (Fafard Growing Mix 2: 70% Canadian sphagnum peat, 30% perlite and vermiculite) that had been autoclaved (25 minutes wet cycle) each of the two days preceding planting for a total of 2 treatments separated by approximately 24 hrs. The autoclaved soil was thoroughly moistened with autoclaved deionized water, prior to filling half- gallon nursery pots. Sterilized seeds were then planted 2 seeds per container,

at approximately one-quarter in. depth, covered with aluminum foil, then refrigerated at 4°C for 5 days.

Plants were randomly assigned positions within a 72 square grid in a Coviron® PGW36 Plant Growth Chamber at the University of Massachusetts Boston. Standard growth conditions were 16 hours of 800 μmol/$m^2$/s intensity light at 23°C and 8 hours dark at 18°C. For the first two weeks in the growth chamber plants were watered 6 days per week with autoclaved deionized water.

Thereafter control plants (C) were watered 2 times per week with autoclaved deionized water and once with autoclaved deionized water supplemented with Peter's 20-20-20 all-purpose fertilizer at a concentration of 120 parts per million. Plants assigned to the nutrient deprived (N) treatment were watered 3 times per week with autoclaved deionized water and no fertilizer. Tissue was collected from three or four biological replicates of *L. sativa* and *L. serriola* in each treatment. In order to minimize variation between samples due to developmental differences, leaf tissue was collected when the first individual flowers of the secondary inflorescence are visible but still closed [145]. Samples were collected at a consistent time of day, between 1 and 2 hours prior to daybreak, to minimize variation in stress-related transcriptomes [94,95]. Two half-inch leaf discs were placed in sterile containers and immediately immersed in liquid nitrogen.

*Reduced representation bisulfite library preparation and sequencing*

Library preparation proceeded as in [146] with the following exceptions. 1.4 $\mu$g of purified DNA was digested for 10 hours with 30 units of BssSI and 30 units of BsoBI. Enzymes were heat inactivated and reactions cleaned-up, prior to end repair, A-tailing

and ligation of methylated adapters according to manufacturer's instructions (NEB, Ipswich, MA). Ligation products were purified and libraries were size selected with a 1.5% Blue Pippin agarose gel cassette for fragment sizes between 250-600 bp. (Sage Science, Beverly, MA) Bisulfite treatment and clean-up were performed as above. Each library was PCR amplified for 13 cycles in a single 50 $\mu$l reaction containing 20 $\mu$ of bisulfite treated sample. Paired-end sequencing (2x100) was performed on a HiSeq 2000 at the University of Massachusetts Boston Center for Personalized Cancer Therapy Genomics Core. Read were trimmed using Trimmomatic and overlapping paired end reads merged as described for whole genome bisulfite sequencing. The bisulfite non-conversion rate was estimated by aligning reads to the *L. sativa* chloroplast genome as described for whole genome bisulfite sequencing.

The genome assemblies of *L. sativa* (v6) and *L. serriola* (v6) were generously provided by the Compositae Genome Project. Sequences were bisulfite converted using Bismark's bismark_genome_preparation. Trimmed reads were aligned to the bisulfite converted and indexed genome using Bismark and bowtie2 using the two step process and counts of methylated and unmethylated reads for each position in the genome were generated as described for whole genome bisulfite sequencing [136]. For comparisons of DMCs between *L. sativa* and *L. serriola*, reads were aligned to the *L. sativa* (v6) genome. For computation of genome-wide methylation levels and DMC detection between samples within a genotype, *L. sativa* reads were aligned the *L. sativa* (v6) genome and *L. serriola* reads were aligned the *L. serriola* (v6) genome.

*Detection of differential methylation*

Only genome positions with at least ten reads in each replicate of each sample, which did not co-localize with known SNPs between *L. sativa* and *L. serriola* and which had greater than zero variance in percent methylation across all replicates were retained for further analysis. Reads were analyzed in R using MethylSig [103]. Local information was included in the estimation of variance but not local methylation level. The differentially methylated cytosines (DMCs) with q-value <0.05 and a methylation difference >= 20% were considered significant. The mRNA and predicted repetitive features which overlapped with these DMC were identified using bedtools intersect.

*Detection of differentially variable methylation between L. sativa and L. serriola*

We utilized the iEVORA algorithm to test the null hypothesis of equal variances of proportion of methylation between biological replicates of *L. sativa* and *L. serriola* using a q-value threshold of 0.001 [104,147]. The predicted protein coding genes and repetitive features which overlapped with these DMC were identified using bedtools intersect.

**Results**

*Relative contribution of environment and genotype to DNA methylation*

*Lactuca sativa* and *L. serriola* differ in their ability to reproduce in disturbed environments. Here we exposed both species to controlled, with fertilizer, conditions (C), and nutrient-stressed, without fertilizer, conditions (N). In the starting parental (S0) generation the effects of nutrient stress were not apparent until approximately one month

after planting, and thus it was not surprising that controlled and nutrient deprived plants of the parental generation did not significantly differ in time to germination or time to development of first through fourth leaves (results not shown). *Lactuca serriola* plants in controlled conditions flowered significantly earlier than *L. sativa* in controlled conditions (Wilcoxon rank sum test, n=3, p-value = 0.0021). *Lactuca sativa* and *L. serriola* plants grown with nutrient stress flowered significantly later than plants grown in controlled conditions (Wilcoxon rank sum test, n = 3, *L. sativa* p-value = 0.0361 and *L. serriola* p-value = 0.0361).

*Lactuca sativa*, but not *L. serriola*, S1 seedlings were significantly affected by the treatment of the parental generation. *Lactuca sativa* seedlings whose parents were nutrient deprived (n=11) had significantly lower above ground wet weight biomass (median = 0.227 g, median absolute deviation (mad) = 0.1586, Wilcoxon rank sum test, p-value = 0.0353, Figure 1A.) and fewer leaves (median = 4 g, mad = 0, p-value = 0.0035; Figure 1B.) than the offspring of non-stressed, controlled parents (n=19). However, these seedling differences associated with parental treatment did not translate to significant differences in time to flowering between the treatments groups (Figure 2).

*Methylation signals in L. serriola and L. sativa in control and no-fertilizer conditions*

To examine if methylation differences associated with growth in nutrient deprived conditions persisted to flowering in *L. sativa* and *L. serriola* we performed reduced representation bisulfite sequencing of two *L. sativa* and two *L. serriola* individuals which had been grown for two generations without fertilizer (NN). These generations and their parental generations were grown concurrently with the plants

sampled for WGBS sequencing and their parental generation [136]. The plants sampled for WGBS were grown in control conditions for two generations (CC) [136]. Hierarchical clustering using the methylation percentages at positions in NN and CC *L. sativa* and *L. serriola* biological replicates generated distinct, high confidence clusters by treatment and library type, with replicates of each genotype forming distinct sub-clades (Figure 3).

We identified DMCs between NN and CC samples of both *L. serriola* and *L. sativa*. In both *L. serriola* and *L. sativa* the median methylation percent across all DMCs was significantly higher in CC relative to NN conditions (Wilcoxon rank sum test, p-values « 0.001). For both *L. serriola* and *L. sativa*, the median methylation percentage of DMCs both within and upstream of protein coding genes was significantly higher in NN samples relative to CC controls (Wilcoxon rank sum test, p-values « 0.001), but significantly lower within annotated repetitive elements in NN samples relative to CC samples (Table 1). In each case the majority of DMC were found in the CHH context.

There were 667 positions which were differentially methylated in both NN *L. sativa* and NN *L. serriola* relative to their conspecific controls, these common DMC represent 28.9% of all unique DMC in these samples (Figure 4). The direction of difference in mean methylation percent between NN and CC samples was consistent for all of the shared DMC's, and 81% of the DMC had higher methylation levels in CC. Genotype-specific DMC represent a larger percentage of DMC within *L. sativa* (63%) than in *L. serriola* (43%). 48 of the common DMC were located within or upstream of 10 protein coding genes; seven of the genes encode proteins of unknown function, one is

annotated as involved in ATP synthesis coupled proton transport, one as having protein serine/threonine activity, one as a structural constituent of ribosome (Table 4).

We identified 47,848 DMCs between *L. serriola* NN and *L. sativa* NN samples, over half (63.6%) of which had higher percent methylation in *L. sativa*. The median methylation percentage was significantly higher in *L. sativa* (80.6%) than in *L. serriola* (52%) (Wilcoxon rank sum test, p-values « 0.001) at the genome level and in all genomic regions analyzed including annotated repetitive elements (Table 1). Though there were many positions (19,843) that were significantly differentially variable between *L. sativa* and *L. serriola* under nutrient deprived conditions, only two of these positions were also differentially variable between *L. sativa* and *L. serriola* in control conditions [136].

*Signals of trans-generational methylation in L. serriola under no-fertilizer conditions*

We performed RRBS of *L. serriola* grown in the following two-generation stress treatments: samples grown in control conditions whose parents were grown in nutrient stress (NC); samples grown in nutrient stress conditions whose parents were also grown in nutrient stress conditions (NN); and samples grown in nutrient stress conditions whose parents were grown in control conditions (CN). WGBS sequencing was previously [136] generated for a fourth treatment in which both parents and offspring were grown under controlled conditions (CC). Three independent biological replicates were generated for each stress treatment. All libraries were aligned to the *L. serriola* genome sequence. The positions covered with at least 10 reads in all WGBS and RRBS samples were retained, returning information on 22,679, 18,585, 73,271 cytosines in the CG, CHG, and CHH contexts respectively. The median methylation percentage for cytosines in the CG context

were 86.96% (mad 11.87%), 94.7% (mad 7.86%), 94.62% (mad 7.97%), 95.12% (mad 7.23%) for plants grown in the CC, NC, CN and NN environments. The CC samples with no history of nutrient stress clustered distinctly from all other samples (Figure 5) though it is not possible with the current data to distinguish biological and technical variation due to the method of sequencing.

We identified DMCs between *L. serriola* NN, CN, and NC samples relative to CC samples. There was no genome-wide statistical difference in median percent methylation of DMCs within protein coding regions detected among the three treatments with a history of nutrient stress (NN, CN, or NC). However, the median methylation percentages showed significant genomic position dependent differences relative to controls. The median percentage methylation of DMC in protein coding genes within 1 kb of an annotated repetitive element was significantly lower in the samples with nutrient deprivation life histories than in controls, but methylation levels at DMC in protein coding genes more than 1 kb from an annotated repetitive element were significantly higher (Wilcoxon rank sum test, $p<0.05$, Table 2).

Within *L. serriola* samples which had themselves been subjected to stress treatment (CN and NN), 4,012 positions were differentially methylated relative to controls and 4,011 these positions had the same direction of change (hypo- or hyper-methylation) relative to controls. These DMC represent 43.9% of all unique DMC in these samples (Figure 7 A.), and 52% of the common DMC had higher methylation levels in CC. 227 of the common DMC were located within or upstream of 57 protein coding genes of unknown function (List 1). Most (76.1%) of the DMC associated with current

stress are also found in NC, and all of the DMC positions had the same direction of difference from controls (Figure 8). Approximately 6% of the common DMC were located within or upstream of 46 protein coding genes of unknown function (List 2). It is also notable that the number of DMC are positively correlated with the severity of stress in terms of immediacy and generational duration; relative to the CC treatment NN samples had the most DMC (7,377), followed by CN (5,765), and NC samples (5,212) (Figure 9).

For all contexts and conditions the coefficient of variation for methylation percentage at a position is inversely related to the average percent methylation, reflecting the importance of high and invariant methylation, possibly in silencing transposable elements (Figure 10). We utilized the iEVORA algorithm to test the null hypothesis of equal variances in the proportions of methylation between biological replicates of stressed and control *L. serriola* using a q-value threshold of 0.001 [104,147]. The methylation levels at differentially variable cytosines (DVC) tended to be high, more than 60% methylation, in all treatment vs. control contrasts (Figure 10 A.). Variability between biological replicates in stress treatments was lower than in controls (Figure 10 B.). However, in the current stress treatments (NN, CN) DVCs tended to be completely methylated and invariable in one of the two treatments (Figure 10). There were approximately equal number of DVCs which were more variable in the two current stress treatments (NN, CN) than control (186 DVCs), and less variable in both current stress treatments relative to untreated controls (203 DVCs), approximately 12 and 14% respectively of all unique DVCs found in both treatments. Methylation

levels at DVC positions in CN samples were more similar to controls and the coefficient of variation took on a range of more similar values (Figure 10 A.).

**Discussion**

The hypermethylation observed in *L. sativa* relative to *L. serriola* grown in controlled conditions [136] was maintained under stress conditions, *L. serriola* was hypomethylated relative to *L. sativa* when comparing between RRBS libraries of *L. sativa* and *L. serriola* grown without fertilizer for two generations (NN). The median level of methylation of DMC positions within NN *L. sativa* samples were consistently and significantly higher than in NN *L. serriola* samples in all genomic regions analyzed including annotated repetitive elements (Table 1). The methylation patterns of *L. sativa* and *L. serriola* NN and CC conditions were more similar by treatment and library type than by genotype (Figure 4).

In both *L. serriola* and *L. sativa*, DMCs in NN samples were globally hypomethylated relative to controls, though methylation levels of DMC differed significantly by genomic region. DMC within protein coding genes and upstream regions were significantly more methylated than in control samples, but significantly less methylated within annotated repetitive elements of NN samples relative to controls (Table 1). Several factors suggest that these significant patterns of hyper- and hypomethylation could be beneficial, targeting higher rates of homologous recombination, and the associated higher rates of mutation, to less deleterious regions of the genome. Increases in homologous recombination frequency have been

reported in plants subjected to abiotic stress and correlated with increased fitness in stressful environments [46]. Recombination hotspots are associated with high mutation frequency [148,149] and, in some plant genomes, enriched in LRR resistance genes [148]. Additionally, the frequency of crossing over events in a region is negatively correlated with its methylation levels as been shown in altered methylation patterning of *met1* mutants [150,151] and reduce recombination at recombination hotspots where constructs target methylation to these regions [150].

The parental generation's nutrient deprivation status resulted in significant differences in the S1 number of leaves and above ground biomass in *L. sativa* but not *L. serriola* (Figure 2). These gross phenotypic differences in the S1 based on parental treatment were not detectable at maturity in plants grown in control conditions, however, we found evidence that treatment specific methylation signals persisted in offspring of nutrient deprived parents grown in control conditions. In contrast, Secco et al. did not see persistence of methylation signals in offspring of inorganic phosphate deficient plants [152], suggesting a possible stress dose dependency of methylation persistence. This is supported by our comparative analysis of *L. serriola* with differing life histories of nutrient deprivation. The number of DMC detected relative to controls were positively correlated with the severity of stress in terms of immediacy and generational duration (Figure 9). This finding highlights the importance of considering the duration of the stress and the time between the application of the stress and when the tissue was sampled when comparing methylation studies.

In *L. serriola* we found apparent conservation of the positions and relative methylation difference of DMC in NN, CN, and NC samples relative to controls. There were 4,012 positions which were differentially methylated in both NN and CN *L. serriola* relative to controls, 43.9% of all DMC in these samples (Figure 7 A.). Most (76%) of these DMC were also differentially methylated in the S1 of stressed parents which had not themselves been subject to stress (NC) (Figure 8). Additional work is required to determine if this apparent conservation represents biological signal or a technical artifact due to comparing RRBS to WGBS libraries. Hierarchical clustering analysis including only the RRBS libraries does not distinguish between treatments (NN, CN, NC) suggests a high degree of relatedness between the biological replicates of different treatments (Figure 11). If confirmed the conservation of these stress associated sites even in samples which had not themselves been subject to stress implies the transgenerational transmission of stress associated differential methylation. Previous studies have suggested that inheritance of methylation may be inconsistent among siblings, though where the inherited methylation signals are present they have been associated with beneficial performance. Using low resolution methylation sensitive amplified polymorphism (MSAP), Kou et al. found offspring of nitrogen deprived plants which showed the altered mC pattern of their parents performed better in nitrogen deprivation than their siblings which did not inherit the modified mC pattern [153]. Likewise MSAP patterns in heavy metal stressed rice showed cases of transmission of the parental modification to the progeny as well as to the next selfed generation, and beneficial performance of offspring of stressed individuals with heavy

metal treatment relative to offspring of non-treated controls [154]. Beneficial effects of adaptive transgenerational priming has been reported to be reduced by treatment with methylation inhibitors [42,46], Dicer loss of function mutants [42,46] and with loss of function Pol IV, the RNA polymerase which produces transcripts from which sRNA are derived [42]. Interestingly, Boyko et al (2010) found offspring of salt stressed *Arabidopsis* were globally hypermethylated relative to offspring of non-stressed controls when grown in control conditions, but were hypomethylated relative to offspring of non-stressed parent when grown on salt [46]. The hypomethylation was also associated with better growth on salt media, but the beneficial effect of parental treatment was reduced when plants were treated with a methylation inhibitor [46]. A possible explanation could that the local hypomethylation is due to the activity of a DNA glycosylase such as ROS1 whose activity is positively regulated by DNA methylation [52]. This mode of action would be complementary with the hypothesis of sRNA of RdDM as the memory mechanism of transgenerational inheritance of modifications [155].

In addition to beneficial plant phenotypes associated with average methylation levels, diversity in methylation has been positively correlated with in plant productivity [75]. We previously found that most positions (86%) having significant differences in variability between *L. sativa* and *L. serriola* had more variable methylation percentages in the domestic variety *L. sativa* [136]. Similarly, increased methylation diversity was seen in domesticated soybean relative to wild soybeans [113]. It is interesting to note that one of the most significant positive measures of productivity observed previously associated with epigenetic diversity was increased plant density, a trait under positive

selection in modern agriculture [156]. Here we see that when both genotypes were grown for two generations in nutrient deprived conditions we found the slight majority (59%) of DVCs were more variable among *L. serriola* replicates. Variability does not appear to be a conserved characteristic of particular genomic loci as the vast majority of positions which were differentially variable between the genotypes in the control conditions [136] were not differentially variable in the nutrient deprivation conditions. This shift in the relative variability of *L. sativa* and *L. serriola* under stressful conditions suggests the possibility that change in variability of methylation in response to changing environment could be a characteristic of stress adaptation.

Our work suggests several roles for acquisition and inheritance of methylation in the evolution of *Lactuca* sp. response to stress. Both genotypes exhibited patterns of hypermethylation within gene bodies and hypomethylation over repetitive elements in treatment conditions relative to conspecific controls, which suggests a beneficial role for stress associated methylation in targeting stress associated higher rates of homologous recombination, and the associated higher rates of mutation, to inter-genic regions of the genome. We also found that changes in relative methylation levels at DMCs are less affected by environment than are changes in relative variability of methylation at DVCs. Though there were significant differences in methylation levels between *L. sativa* and *L. serriola*, in both treatment and controlled conditions most DMC were hypermethylated at DMC in *L. sativa* relative to *L. serriola* and there were a significant number of DMC in nutrient deprived conditions that were found in both genotypes, with the same direction of difference, relative to conspecific controls. In

contrast, the frequency of DVC which were more variable in *L. sativa* relative to *L. serriola,* shifted between controlled and nutrient stressed conditions and there was very little overlap between DVC positions in the genotypes relative to controls. We found suggestions that abiotic stress associated methylation may be transmitted between generations with fidelity.

## Figures

**Figure 1.** Development characteristics of *L. sativa* and *L. serriola* in one month old seedlings by parental treatment. The offspring of a parental generation grown in either nutrient deprived (N) or controlled (C) conditions had genotype dependent differences in wet weight (A.) and number of leaves (B.) as one month old seedlings. Dots represent individual seedlings and bars represent the median value per treatment.

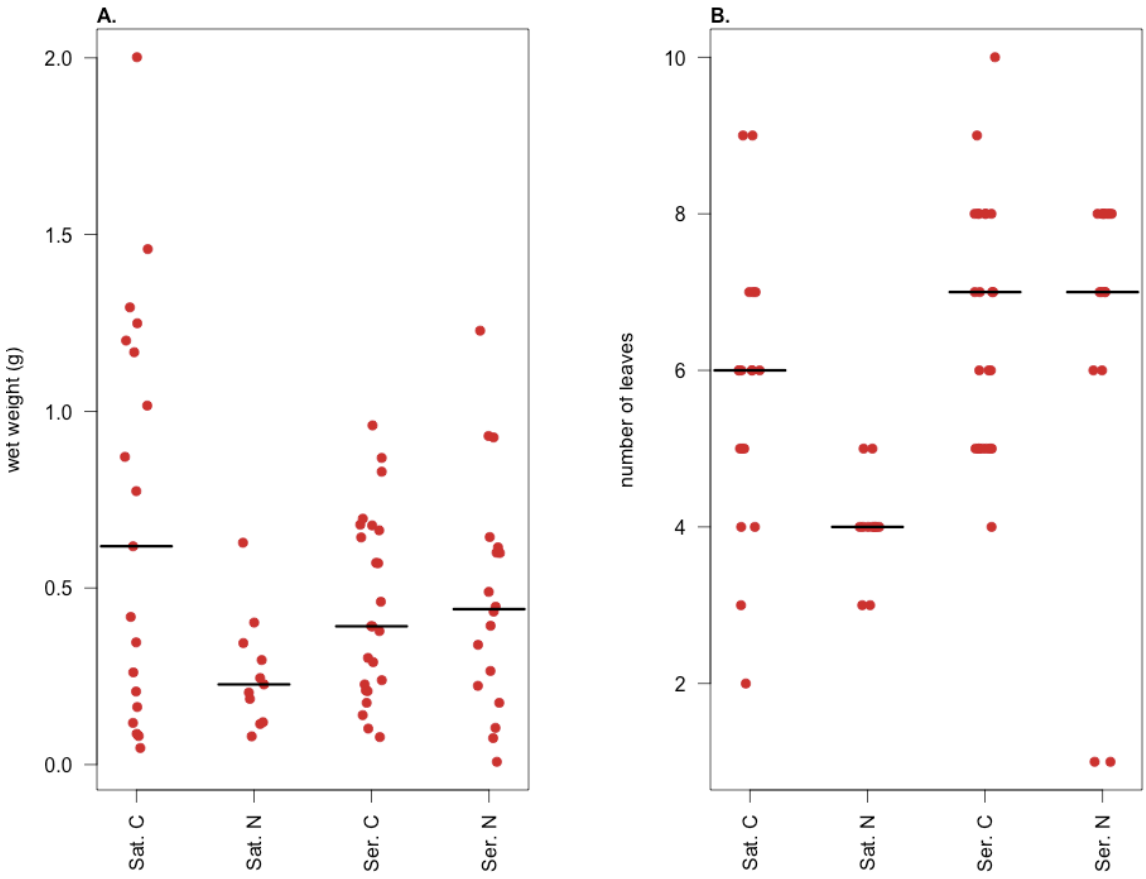


**Figure 2.** Effect of parental stress treatment on days to flowering of next generation in *L. sativa* and *L. serriola* by parental treatment. Days to flowering for *L. sativa* and *L. serriola*, where parental and offspring generations were grown in control conditions (CC), parental generation in control conditions offspring generation in treatment (CN), parental generation in treatment conditions offspring generation in control conditions (NC), and parental and offspring generations were grown in treatment conditions (NN).

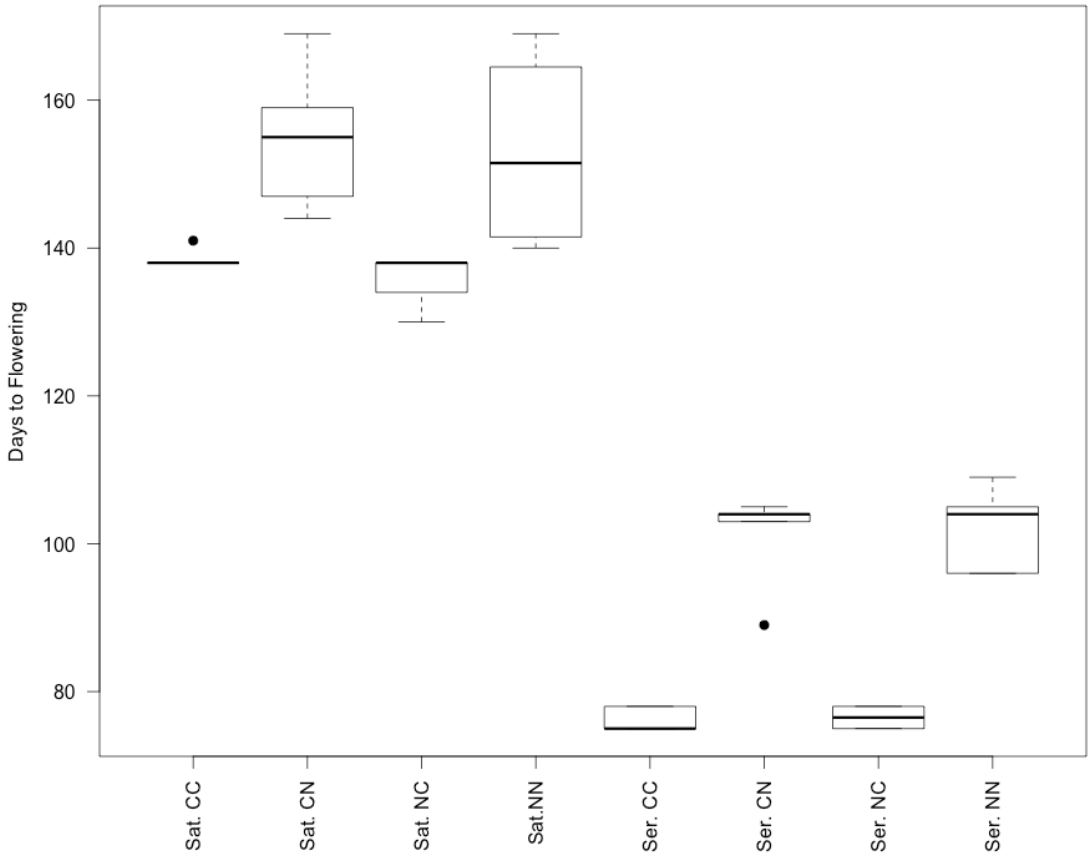


**Figure 3.** Hierarchical clustering of methylation of *L. sativa* and *L. serriola* grown in different treatment conditions. Shown are hierarchical clustering of methylation at positions having sufficient read support in CG (A), CHG (B), and CHH contexts. In red are approximately unbiased probability values generated by R package pvclust, in green are boot strap probabilities. C1 and C2 represent biological replicates grown in control conditions, N1 and N2 represent biological replicates grown in no-fertilizer conditions.

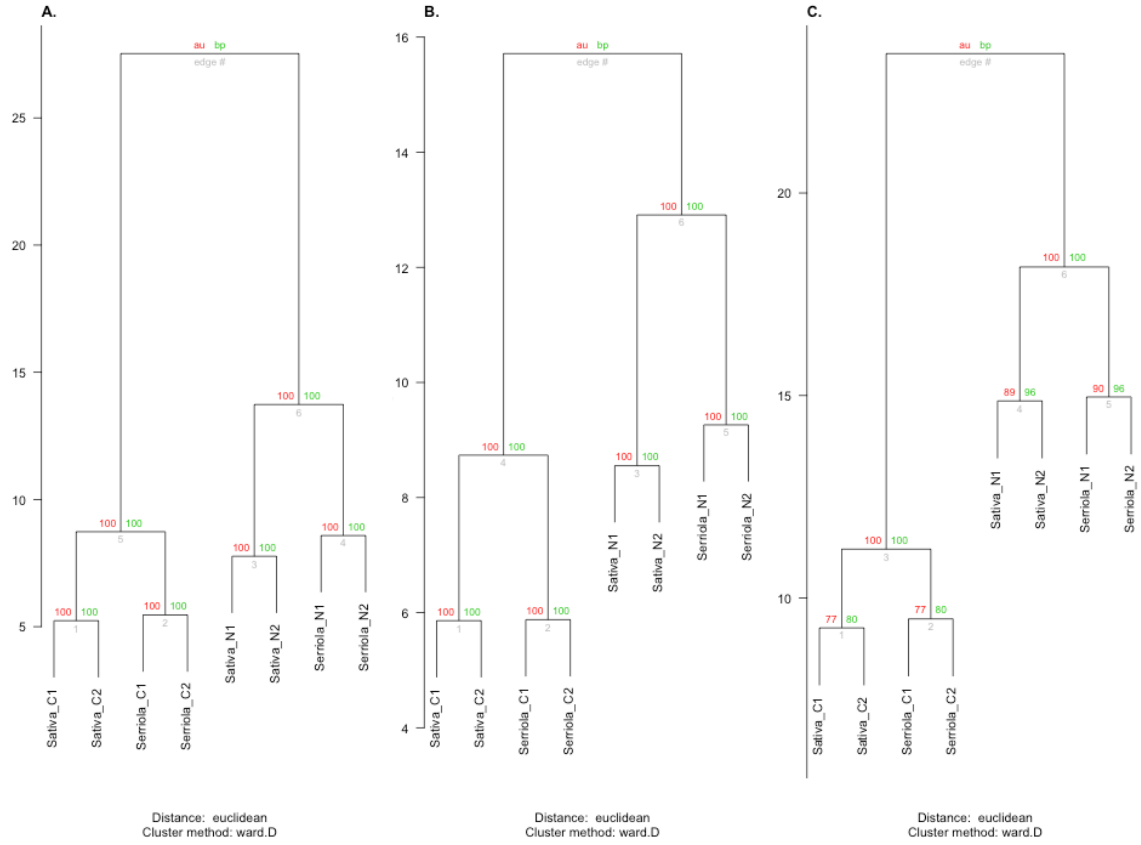


**Figure 4.** Venn diagram of DMCs found in NN *L. sativa* and NN *L. serriola* relative to their conspecific controls.

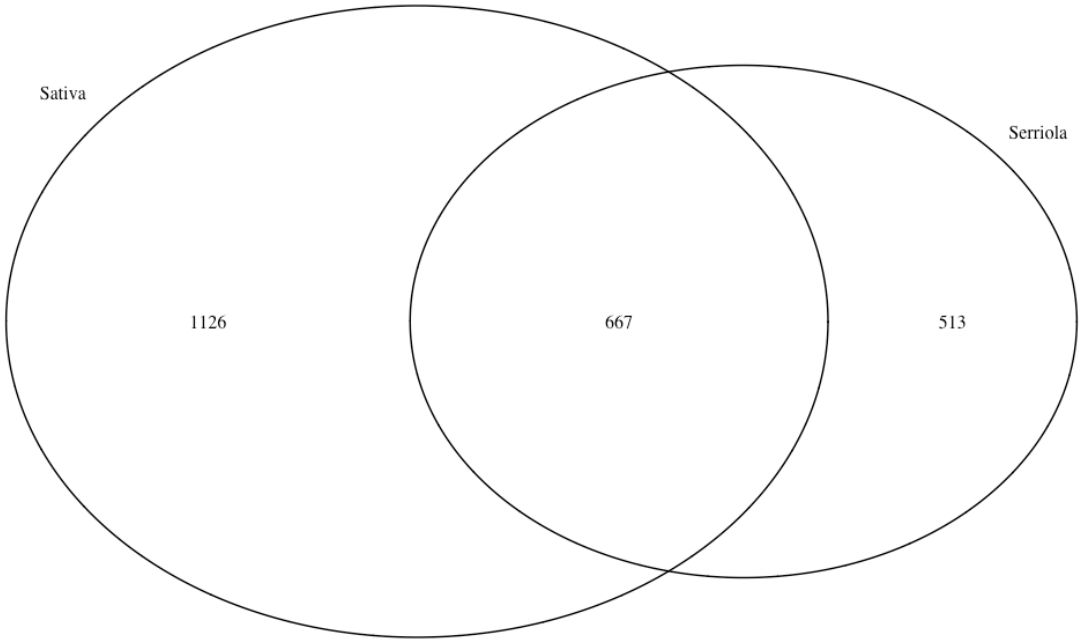


**Figure 5.** Hierarchical clustering of methylation at positions having sufficient read support in CG (A), CHG (B), and CHH (C) contexts. In red are approximately unbiased probability values generated by R package pvclust, in green are boot strap probabilities. CC1, CC2, and CC3 represent biological replicates grown in control conditions having parents also grown in control conditions, CN1, CN2, and CN3 represent biological replicates grown in without fertilizer and parental plants grown in control conditions, NC1, NC2, and NC3 represent biological replicates in control conditions whose parental plants were grown in no-fertilizer conditions, NN1, NN2, and NN3 represent biological replicates grown without fertilizer for two consecutive generations.

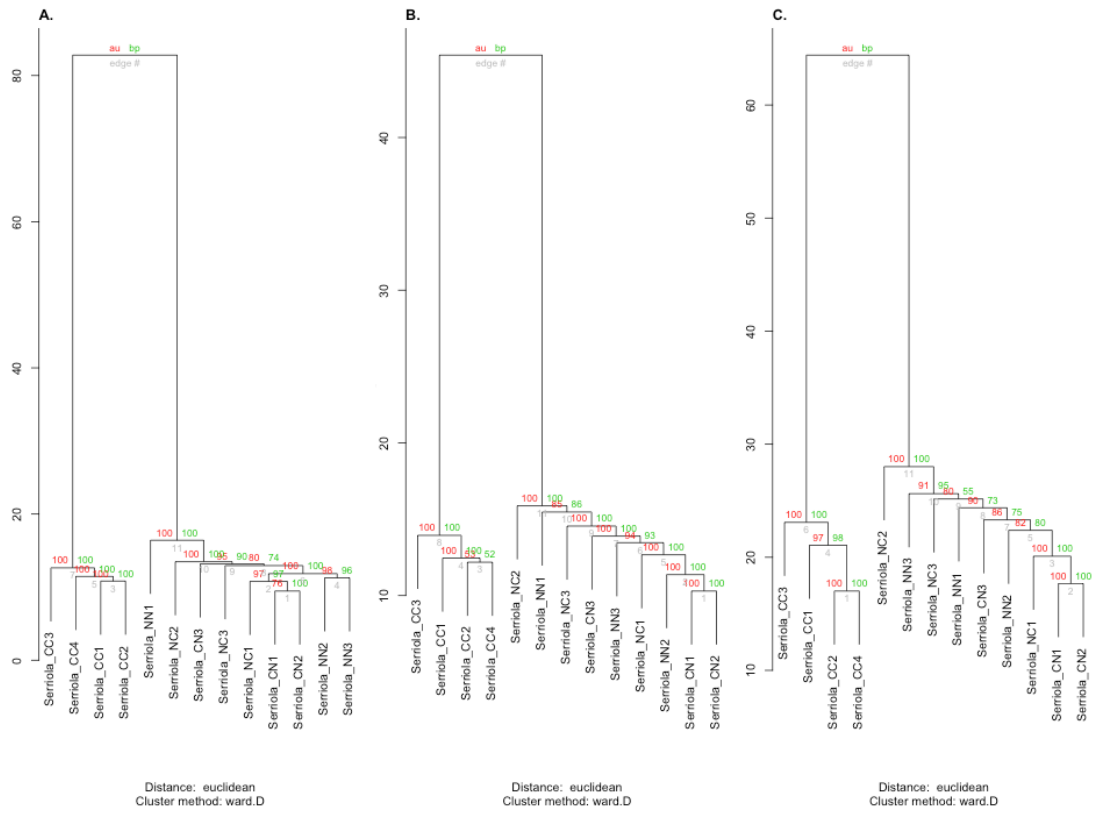


**Figure 6.** Total number of DMCs in *L. serriola* with different family histories of nutrient deprivation. The graph shows the total number of DMCs detected in *L. serriola* grown in nutrient deprived conditions for two generations (NN), *L. serriola* grown in nutrient deprived conditions and offspring of parents grown in controlled conditions (CN), or *L. serriola* grown in control conditions and *L. serriola* offspring of nutrient deprived parents (NC). 76% of the positions which were differentially methylated in both of the current stress treatments (NN, CN) are also differentially methylated in unstressed offspring of stressed parents (NC).

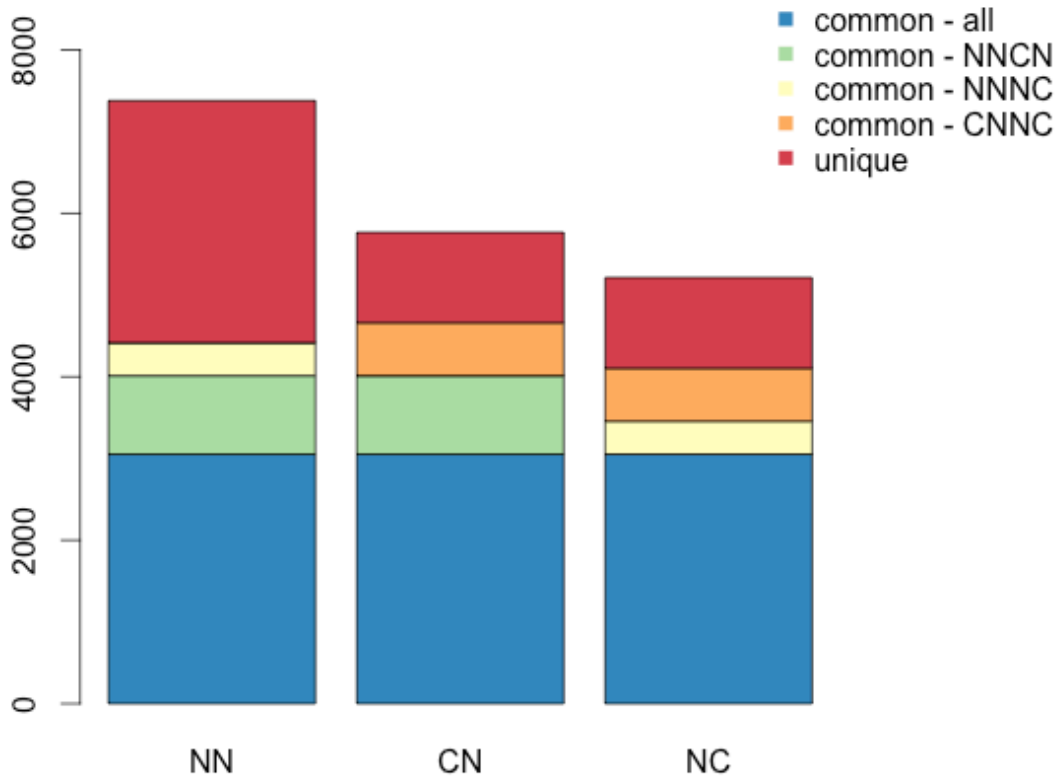


**Figure 7.** Relative methylation levels in nutrient deprived and control *L. sativa* and *L. serriola* plants. In both *L. sativa* (A.) and *L. serriola* (B.) DMCs in nutrient deprived individuals (NN) were hypomethylated relative to controls (CC).

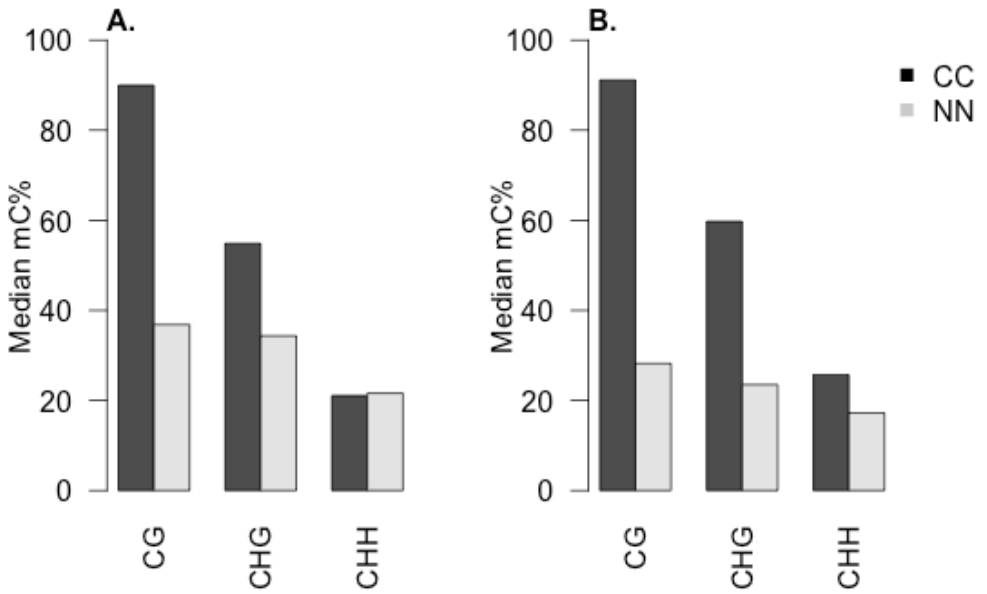


**Figure 8.** Relative methylation levels in nutrient deprived *L. sativa* and *L. serriola* by genomic region. DMCs between NN *L. sativa* and NN *L. serriola* were consistently hypomethylated in *L. serriola*.

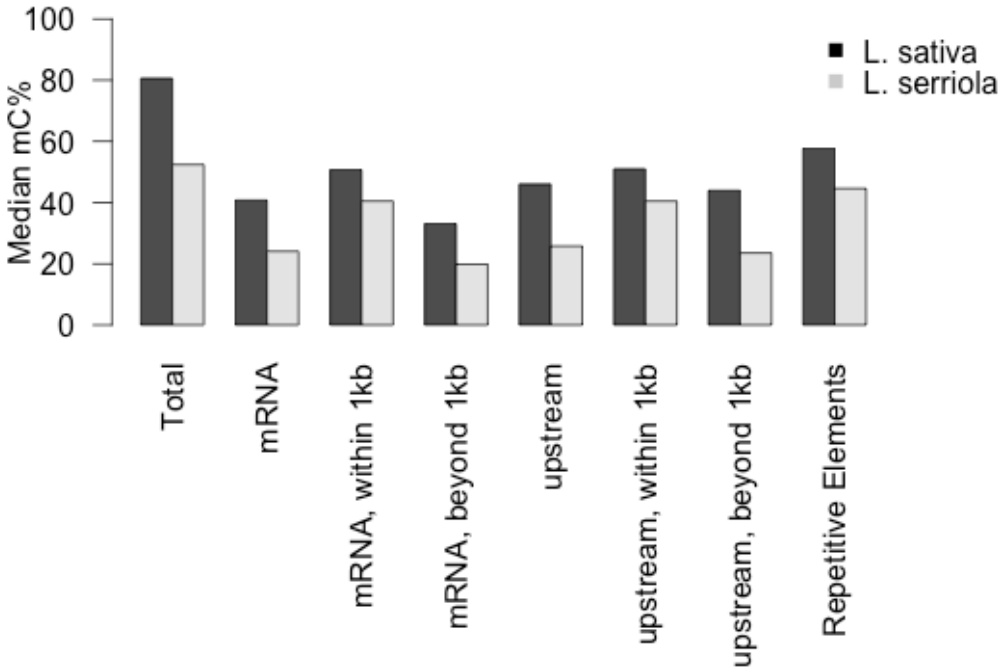


**Figure 9.** Relationship of the variance in methylation and mean methylation in *L. serriola* at cytosines covered by more than 10 reads in all samples.

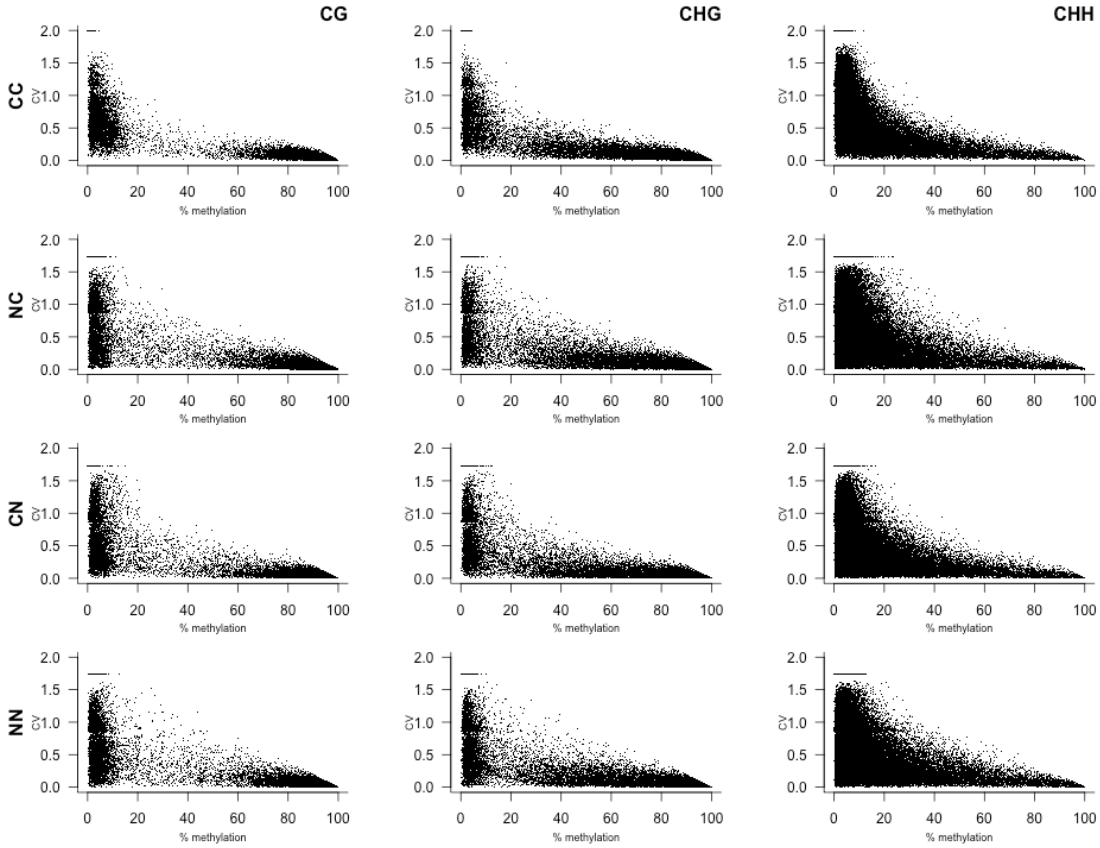


**Figure 10.** Relationship of average methylation levels and variance in methylation in *L. serriola* with different family histories of nutrient deprivation. Average methylation levels (A) and variance in methylation (B) are shown for NC, CN and NN treatments in *L. serriola*.

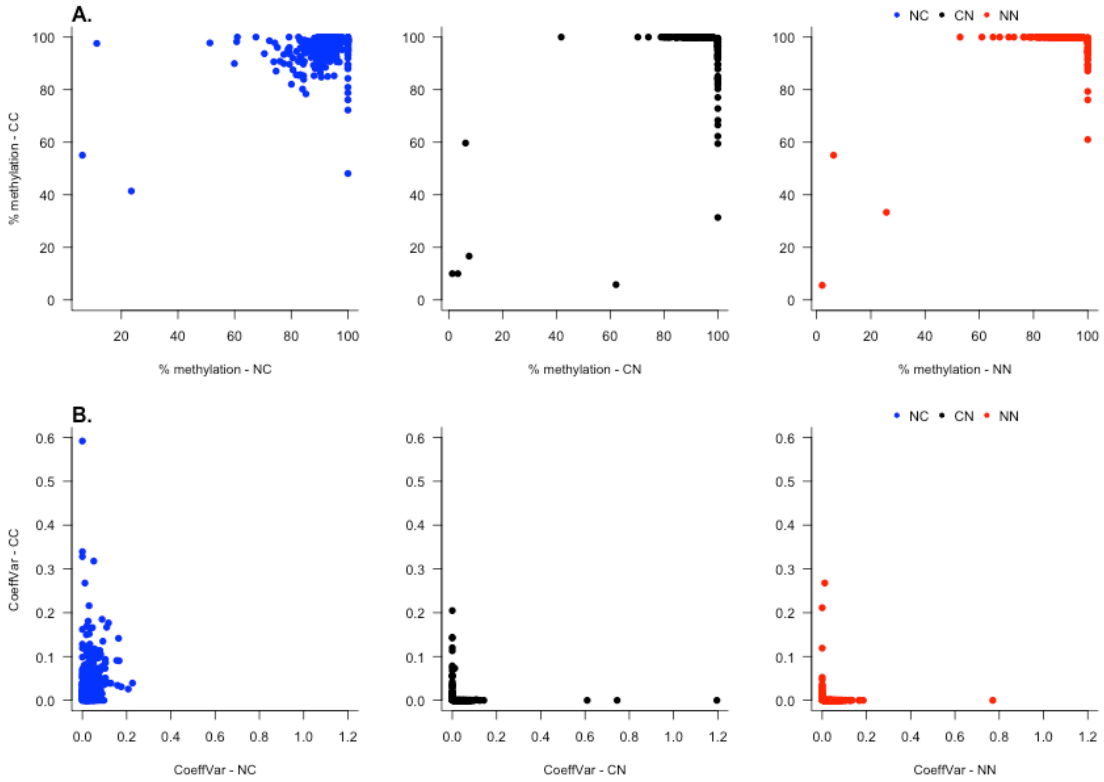
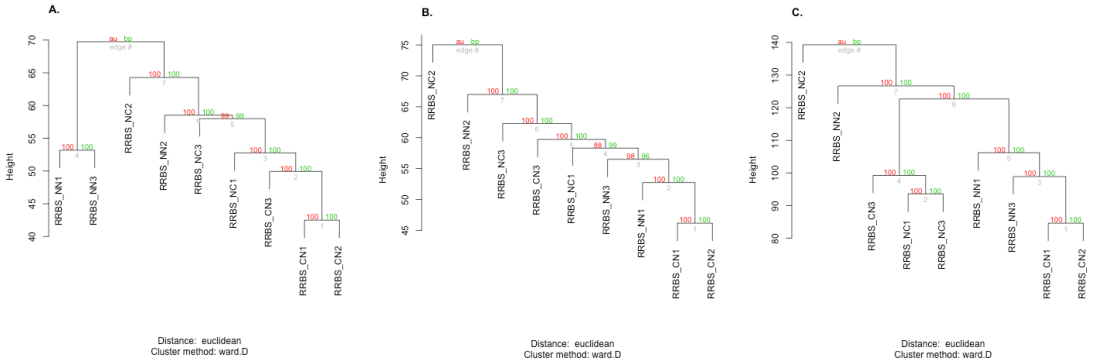


**Figure 11.** Hierarchical clustering of methylation at positions having sufficient read support in CG (A), CHG (B), and CHH (C) contexts. CC1, CC2, and CC3 represent biological replicates grown in control conditions having parents also grown in control conditions, CN1, CN2, and CN3 represent biological replicates grown in without fertilizer and parental plants grown in control conditions, NC1, NC2, and NC3 represent biological replicates in control conditions whose parental plants were grown in no-fertilizer conditions, NN1, NN2, and NN3 represent biological replicates grown without fertilizer for two consecutive generations.



## Tables

**Table 1.** Comparison of median methylation percentages over genomic regions of *L. sativa* and *L. serriola* in NN and CC conditions. Each comparison considers only positions with at least 10 reads in both samples.

| | median, NN sativa | median, CC sativa | p-value | median, NN serriola | median, CC serriola | p-value | median, NN sativa | median, NN serriola | p-value |
|---|---|---|---|---|---|---|---|---|---|
| mRNA | 24.23 | 1.15 | 0.00 | 30.46 | 3.06 | 0.00 | 40.75 | 24.07 | 0.06 |
| mRNA, within 1 kb | 21.43 | 0.73 | 0.00 | NA | NA | NA | 50.70 | 40.48 | 0.00 |
| mRNA, beyond 1 kb | 24.23 | 1.20 | 0.00 | 28.43 | 3.01 | 0.00 | 33.08 | 19.87 | 0.00 |
| Upstream | 26.53 | 1.58 | 0.00 | 42.30 | 3.34 | 0.00 | 46.05 | 25.76 | 0.00 |
| Upstream, within 1 kb | 21.43 | 0.73 | 0.00 | NA | NA | NA | 51.01 | 40.48 | 0.00 |
| Upstream, beyond 1 kb | 27.14 | 1.65 | 0.00 | 42.00 | 3.34 | 0.00 | 43.94 | 23.53 | 0.00 |
| Repetitive Elements | 7.18 | 65.62 | 0.00 | 5.99 | 60.00 | 0.00 | 57.69 | 44.61 | 0.00 |

**Table 2.** Comparison of median methylation percentages over genomic regions of *L. serriola* in NN, CN and NC and CC conditions. Each comparison considers only positions with at least 10 reads in both samples.

| | median, NN serriola | median, CC serriola | p-value | median, CN serriola | median, CC serriola | p-value | median, NC serriola | median, CC serriola | p-value |
|---|---|---|---|---|---|---|---|---|---|
| mRNA | 29.31 | 38.93 | 0.92 | 31.97 | 40.09 | 0.52 | 30.95 | 46.39 | 0.02 |
| mRNA, within 1 kb | 31.04 | 46.39 | 0.03 | 35.53 | 46.84 | 0.00 | 34.12 | 49.77 | 0.00 |
| mRNA, beyond 1 kb | 21.86 | 2.26 | 0.00 | 25.00 | 2.92 | 0.00 | 22.47 | 6.56 | 0.03 |
| Upstream | 39.89 | 21.15 | 0.14 | 41.55 | 24.71 | 0.49 | 23.05 | 18.75 | 0.21 |
| Upstream, within 1 kb | 38.28 | 40.90 | 0.61 | NA | NA | NA | 27.16 | 59.15 | 0.70 |
| Upstream, beyond 1 kb | 40.54 | 19.55 | 0.12 | 41.92 | 24.66 | 0.46 | 23.05 | 18.57 | 0.25 |
| Repetitive Elements | 24.62 | 26.88 | 0.06 | 22.98 | 49.99 | 0.00 | 23.85 | 49.18 | 0.00 |

**Table 3.** Comparison of median methylation percentages over genomic regions of *L. serriola* in different conditions, by proximity to annotated repetitive elements. Each comparison considers only positions with at least 10 reads in both samples.

| | mRNA, within 1 kb | mRNA, beyond 1 kb | p-value | upstream, within 1 kb | upstream, beyond 1 kb | p-value |
|---|---|---|---|---|---|---|
| NN serriola (NNCCser) | 31.04 | 21.86 | 0.29 | 38.28 | 40.54 | 0.67 |
| CC serriola (NNCCser) | 46.39 | 2.26 | 0.00 | 40.90 | 19.55 | 0.23 |
| CN serriola (CNCCser) | 35.53 | 25.00 | 0.28 | 15.79 | 41.92 | 0.38 |
| CC serriola (CNCCser) | 46.84 | 2.92 | 0.00 | 44.13 | 24.66 | 0.80 |
| NC serriola (NCCCser) | 34.12 | 22.47 | 0.73 | 27.16 | 23.05 | 0.97 |
| CC serriola (NCCCser) | 49.77 | 6.56 | 0.00 | 59.15 | 18.57 | 0.30 |

**Table 4.** Protein coding genes containing DMC within upstream or within gene bodies that were found in both NN *L. sativa* and NN *L. serriola* relative to their conspecific controls.

| mRNA ID | Gene ontology terms |
|---|---|
| Lsat_1_v5_gn_1_24401 | |
| Lsat_1_v5_gn_3_17360 | |
| Lsat_1_v5_gn_3_17381 | GO:0015078: hydrogen ion transmembrane transporter activity: Molecular Function \| GO:0015986: ATP synthesis coupled proton transport: Biological Process \| GO:0015991: ATP hydrolysis coupled proton transport: Biological Process \| GO:0033177: proton-transporting two-sector ATPase complex |
| Lsat_1_v5_gn_3_22600 | GO:0003735: structural constituent of ribosome: Molecular Function \| GO:0005622: intracellular: Cellular Component \| GO:0005840: ribosome: Cellular Component |
| Lsat_1_v5_gn_4_102681 | |
| Lsat_1_v5_gn_4_145480 | |
| Lsat_1_v5_gn_4_152881 | GO:0004672: protein kinase activity: Molecular Function \| GO:0004674: protein serine/threonine kinase activity: Molecular Function \| GO:0005515: protein binding: Molecular Function \| GO:0005524: ATP binding: Molecular Function |
| Lsat_1_v5_gn_4_63021 | |
| Lsat_1_v5_gn_4_65980 | |
| Lsat_1_v5_gn_6_37840 | |

## Lists

**List 1.** Identifiers of 57 protein coding genes of unknown function which were associated with DMC found in *L. serriola* samples subjected to stress.

| | | |
|---|---|---|
| Lser_1_v1_gn_1_57860 | Lser_1_v1_gn_8_46860 | Lser_1_v1_gn_4_57800 |
| Lser_1_v1_gn_1_67861 | Lser_1_v1_gn_8_6600 | Lser_1_v1_gn_4_71541 |
| Lser_1_v1_gn_1_34840 | Lser_1_v1_gn_8_58961 | Lser_1_v1_gn_4_76341 |
| Lser_1_v1_gn_2_160 | Lser_1_v1_gn_8_65781 | Lser_1_v1_gn_4_82241 |
| Lser_1_v1_gn_2_36280 | Lser_1_v1_gn_8_73660 | Lser_1_v1_gn_4_98880 |
| Lser_1_v1_gn_2_4440 | Lser_1_v1_gn_8_83821 | Lser_1_v1_gn_5_33080 |
| Lser_1_v1_gn_2_5240 | Lser_1_v1_gn_8_23320 | Lser_1_v1_gn_5_64780 |
| Lser_1_v1_gn_2_17381 | Lser_1_v1_gn_8_30321 | Lser_1_v1_gn_6_43921 |
| Lser_1_v1_gn_3_35600 | Lser_1_v1_gn_9_32460 | Lser_1_v1_gn_6_45761 |
| Lser_1_v1_gn_3_760 | Lser_1_v1_gn_1_43960 | Lser_1_v1_gn_6_27020 |
| Lser_1_v1_gn_3_65721 | Lser_1_v1_gn_1_59160 | Lser_1_v1_gn_7_57641 |
| Lser_1_v1_gn_3_9461 | Lser_1_v1_gn_2_21861 | Lser_1_v1_gn_7_23081 |
| Lser_1_v1_gn_3_9481 | Lser_1_v1_gn_2_5221 | Lser_1_v1_gn_7_31001 |
| Lser_1_v1_gn_4_43060 | Lser_1_v1_gn_2_1800 | Lser_1_v1_gn_8_44021 |
| Lser_1_v1_gn_4_16001 | Lser_1_v1_gn_2_13661 | Lser_1_v1_gn_8_6001 |
| Lser_1_v1_gn_5_40201 | Lser_1_v1_gn_3_55720 | Lser_1_v1_gn_8_6020 |
| Lser_1_v1_gn_5_47480 | Lser_1_v1_gn_3_63980 | Lser_1_v1_gn_8_70301 |
| Lser_1_v1_gn_7_38501 | Lser_1_v1_gn_3_67420 | Lser_1_v1_gn_8_10340 |
| Lser_1_v1_gn_7_44920 | Lser_1_v1_gn_3_1920 | Lser_1_v1_gn_9_55720 |

**List 2.** Identifiers of 48 protein coding genes of unknown function which were associated with DMC found in all *L. serriola* samples either presently subjected to nutrient deprivation or whose parents were subjected to nutrient deprivation.

| | | |
|---|---|---|
| Lser_1_v1_gn_1_57860 | Lser_1_v1_gn_7_44920 | Lser_1_v1_gn_4_76341 |
| Lser_1_v1_gn_1_67861 | Lser_1_v1_gn_8_46860 | Lser_1_v1_gn_4_98880 |
| Lser_1_v1_gn_1_34840 | Lser_1_v1_gn_8_58961 | Lser_1_v1_gn_5_33080 |
| Lser_1_v1_gn_2_160 | Lser_1_v1_gn_8_83821 | Lser_1_v1_gn_5_64780 |
| Lser_1_v1_gn_2_36280 | Lser_1_v1_gn_8_23320 | Lser_1_v1_gn_6_43921 |
| Lser_1_v1_gn_2_17381 | Lser_1_v1_gn_8_30321 | Lser_1_v1_gn_6_45761 |
| Lser_1_v1_gn_3_35600 | Lser_1_v1_gn_9_32460 | Lser_1_v1_gn_6_27020 |
| Lser_1_v1_gn_3_760 | Lser_1_v1_gn_1_59160 | Lser_1_v1_gn_7_57641 |
| Lser_1_v1_gn_3_65721 | Lser_1_v1_gn_2_21861 | Lser_1_v1_gn_7_23081 |
| Lser_1_v1_gn_3_9461 | Lser_1_v1_gn_2_13661 | Lser_1_v1_gn_7_31001 |
| Lser_1_v1_gn_3_9481 | Lser_1_v1_gn_3_55720 | Lser_1_v1_gn_8_44021 |
| Lser_1_v1_gn_4_43060 | Lser_1_v1_gn_3_63980 | Lser_1_v1_gn_8_70301 |
| Lser_1_v1_gn_4_16001 | Lser_1_v1_gn_3_67420 | Lser_1_v1_gn_8_10340 |
| Lser_1_v1_gn_5_40201 | Lser_1_v1_gn_3_1920 | Lser_1_v1_gn_9_55720 |
| Lser_1_v1_gn_5_47480 | Lser_1_v1_gn_4_57800 | |
| Lser_1_v1_gn_7_38501 | Lser_1_v1_gn_4_71541 | |

# CHAPTER 5

# CONCLUSION

The research presented in this dissertation increases our understanding of the relationship between DNA methylation in plants and environmental conditions through bisulfite sequencing of closely related accessions of *Lactuca*. We have shown that the methylomes of domesticated *L. sativa* and its conspecific wild and weedy relative, *L. serriola,* are very similar at the genome scale under non-stressed conditions. Both the domesticated and wild genotypes have genomic-region specific patterns of hypo- and hyper-methylation under stress conditions which may direct stress associated increases in homologous recombination away from gene coding regions. Both genotypes also have conserved methylation signatures around pathogen response related genes when grown in unstressed control conditions. The genotypes also showed environment specific differences in the relative abundance of differentially variable positions, suggesting genotype specific interactions between variability in DNA methylation and environmental stress. Together these findings suggest that epigenetic modifications are an additional source and mechanism of genomic variation which may be isolated and adapted for improvement of crops.

**Relationship of global methylation levels and gene regions**

Though both *L. sativa* and *L. serriola* both have very high average levels of methylation, similar to other large plant genomes, a significant number of DMCs are more highly methylated in *L. sativa* than *L. serriola*. The differences between this self-crossing dicot wild-domestic pair do not support an obvious relationship between methylation levels and a plant's domestication status that transcends diverse plant families. Methylation levels between closely related wild and domesticated monocots are not correlated with domestication status [54] and methylation levels at DMCs between domesticated corn and wild teosinte tend to be less methylated in the domestic varieties [157].

Within the methylomes of *L. sativa* and *L. serriola* are interesting characteristics suggesting a role for DNA methylation in plant-microbe interactions. We found striking patterns of methylation around resistance genes, highly conserved methylation states in pathogen response related genes, and gene ontology enrichment of plant-microbe interaction related terms among genes with differentially methylated cytosines between *L. sativa* and *L. serriola*. DMCs that distinguish these genotypes are related to important gene ontology categories known to affect fitness. DMCs between *L. sativa* and *L. serriola* were found in genes enriched for gene ontology terms related to: photosynthesis, biosynthesis of aromatic amino acids, signaling and lipid signaling, and transport; these terms are also implicated in plant microbe interactions [158–160].

Methylation levels around protein coding genes in *Lactuca* sp. have characteristic patterns shared by most angiosperms and conserved and striking methylation patterns

around genes related to pathogen response. Like most other angiosperms, methylation levels upstream and downstream of protein coding genes in *Lactuca* sp. are relatively high, dipping dramatically at transcription start and end sites. However, the average levels of methylation in *Lactuca* sp. take on strikingly different patterns around annotated disease resistance genes. The up- and downstream regions of resistance genes have much higher average levels of methylation in the CHH context than the genome wide average across all predicted protein coding genes. The accumulation of CHH methylation is particularly noteworthy in the regions 100-400 bp upstream of the transcription start site. Small RNAs were found to accumulated in these regions of resistance genes in Arabidopsis, and the active demethylation of these regions by ROS1 upon pathogen challenge was shown to be required for pathogen resistance [41]. As the average levels of methylation over these regions do not vary between unchallenged *L. sativa* and *L. serriola*, genotypes with divergent pathogen resistance phenotypes, it would be interesting to compare the relative rates of active methylation and demethylation around resistance genes in *L. sativa* and *L. serriola* during the course of active pathogen infection. Previous work profiling the DNA methylation of salt resistant and tolerant rice genotypes found that rates of change in methylation, but not average methylation levels, differed between salt tolerant and sensitive rice varieties [45].

An additional distinctive feature of *Lactuca* sp. resistance genes is the strong spike in methylation levels in all sequence contexts found at the 3' end of the genes. Similar spikes in methylation had previously been seen in Arabidopsis resistance gene RPP7 [161]. In RPP7 the 3' methylation was associated with an intronic transposable

element, accumulation of Histone H3 lysine9 di-methylation (H3K9me2) and alternative polyadenylation of the gene [161]. Sequence analysis of the 3' intronic regions of these genes in *Lactuca* sp. could determine if similar transposon "domestication" explains the observed accumulation of 3' methylation in lettuce. Likewise, transcripts of these genes could be analyzed for isoform production and temporal changes in methylation levels under pathogen challenge. Though the *L. sativa* and *L. serriola* average levels of methylation around resistance genes do not significantly differ, analysis of differential rates of methylation and demethylation of resistance genes and flanking regions could highlight differences in pathogen responsiveness which could complement traditional gene based crop improvement strategies.

The relationship between average methylation levels and the variability of methylation levels between biological replicates was highly dependent on genomic region. At most genomic positions average methylation levels are inversely related to the variability of methylation at that position. Sites of highly conserved methylation were defined as being among the 25% least variable positions between biological replicates of both *L. sativa* and *L. serriola*. Genes with highly conserved methylation states were highly methylated, even though, genome wide, most positions with highly conserved states had low or no methylation. This seemingly contrary result was explained by looking at the frequency of occurrence of conserved methylation states by genomic region. The majority of positions with highly conserved methylation states were located in unannotated intergenic regions and had extremely low levels of methylation in all sequence contexts. Most of the conserved positions found in annotated repetitive regions

were in the CG and CHG context and had high levels of methylation in line with genome wide averages for these features. Genes containing positions with highly conserved methylation states had unusually high levels of methylation (>90% methylation in all sequence contexts), particularly notable in the CHH context, suggesting that these sites may be targets of RdDM. Annotated genes with the highest frequency of positions with highly conserved methylation states controlling for variation in gene length included endo-1,4-beta-mannosidase, a cell wall degrading enzyme, Flagellin-sensitive 2 (FLS2), an important sensor of pathogen attack, and F-type H+-transporting ATPase subunit C, a conserved transport ATPase. The frequency of conserved methylation states within FLS2 is particularly interesting given previous work correlating expression of FLS2 in ROS1 dependent DNA de-methylation [41]. FLS2 is strongly induced by the bacterial flagellin N-terminal epitope flg22, and triggers transcriptional regulation of stress response genes. Flg22 exposure also results in ROS1 dependent DNA de-methylation and expression of a specific subset of transposable elements [41]. Given the highly conserved and highly methylated state of FLS2 in non-pathogen challenged *Lactuca* sp. and the interaction of FLS2 with de-methylation dependent pathogen resistance – it could be illuminating to assess the methylation in this gene and the class of R-genes targeted by ROS1 over the course of pathogen infection.

The effect of altered methylation levels on plant-microbe interactions and microbial community composition warrants further research. The work to date has focused on the interaction of particular loss of function DNA methylation mutants on the plants' susceptibility to particular pathogens [40,41,62,69,84] or the plant methylation

signature of genotype specific interactions with a particular beneficial microbe [162]. The natural milieu in which the microbes and plant communicates have evolved is much more complex than the introduction of a single beneficial or pathogenic bacteria or fungi in a controlled environment. The plant microbiome can be an important contributor to plant fitness [163] and may be an important link in the interaction of epigenetic diversity of plant populations and their environments with fitness traits such as biomass density, competition and pathogen resistance [75].

**Acquisition and variability of mC in *L. sativa* and *L. serriola* in nutrient limited conditions**

Methylation levels in both nutrient deprived and control samples of *L. sativa* were higher than the corresponding treatment in *L. serriola*. Though *L. sativa* and *L. serriola* differed in methylation levels, there were consistent patterns of methylation between the two conditions. More than half of the DMCs in stressed *L. serriola* relative to controls were also found in *L. sativa* relative to controls with the same direction of difference relative to controls. In both *Lactuca* sp. the plants grown in nutrient limited conditions were hypermethylated in gene bodies and hypomethylated over repetitive elements relative to control plants. In both biotic and abiotic stress response, differential methylation has been associated with genes and upstream regions that are in close proximity to repetitive elements [40,70]. A similar general role for TE associated DNA methylation of nearby genes is found in *Lactuca* as the methylation levels of genes in both *L. sativa* and *L. serriola* differ based on their proximity to annotated repetitive

elements. Methylation levels are significantly higher in genes located within 1000 bp of an annotated repetitive element, possibly due to TE targeted methylation which may introduce selectable variation in gene expression.

We found a significant correspondence between positions of stress associated differences in mean methylation levels between genotypes, but a significant shift in the positions and relative abundance of differentially variable cytosines between the two genotypes under stress conditions. Growth in nutrient limited conditions resulted in accumulation of more DMCs within *L. sativa* relative to con-specific controls than *L. serriola*, suggesting the domestic methylome was more affected by growth in nutrient limited conditions. Though the total number of positions differed dramatically, more than half of the DMC seen in stressed *L. serriola* relative to controls are also seen in *L. sativa* relative to controls with the same direction of difference relative to controls. In contrast, the relative variability at sites of differentially variable methylation shifted between the two genotypes in control and nutrient deprived treatments. Most DVCs were more variable in *L. sativa* than *L. serriola* under control condition but were less variable in *L. sativa* under nutrient limited conditions. The relative shift toward more variability under nutrient limited conditions in *L. serriola* is an interesting finding in light of the relative adaptation of *L. serriola* to marginalized and highly disturbed environments. Recent work identified an association of increased diversity of methylation with increased plant productivity in pathogen and competition challenged environments [75]. The drivers of differential variability are unclear, though variability does not appear to be characteristic of particular genomic regions as sites of differentially variable methylation between

biological replicates of *L. sativa* and *L. serriola* in control conditions were conserved in nutrient limited conditions. It would be interesting to study the relative activity and fidelity of the methyltransferases in the two genotypes to further dissect possible sources of variability.

**Conclusions**

This work identified several ways in which differential methylation levels and conservation of methylation states between these two genotypes are associated with gene regions and functions implicated in plant-microbe interactions. Thought there were particular positions having significantly different average methylation levels, there were also positions in genes related to plant microbe interactions whose methylation states were highly conserved. These instances suggest the value of closely monitoring these loci over the time course of acute stress, particularly over the time course of pathogen exposure and infection. This work also identified an interaction between stress environment and the relative frequency of differentially variable cytosines between the genotypes. This work allows future researchers to be more targeted in their approach to dissect the role of DNA methylation in the stress adaptive phenotypes of these two *Lactuca* sp.. Research may be targeted to key genes and genomic regions involved in pathogen response and more efficiently analyze the time course component of acquired DNA methylation modifications. Additionally the increase in relative abundance of highly variable positions in stress conditions within the relatively stress-adapted genotype, is in line within initial research in Arabidopsis suggesting a positive adaptive

role for increased epigenetic diversity in the absence of genetic diversity [75]. Competition experiments between populations of differing levels of epigenetic diversity within genetically homogeneous populations of *L. sativa* and *L. serriola* would help distinguish the relative contribution of epigenetic diversity and genotype to the stress-adaptation of the wild species.

REFERENCE LIST


1. Smykal P, K Varshney R, K Singh V, Coyne CJ, Domoney C, Kejnovsky E, et al. From Mendel's discovery on pea to today's plant genetics and breeding : Commemorating the 150th anniversary of the reading of Mendel's discovery. Theor Appl Genet. Germany; 2016;129: 2267–2280. doi:10.1007/s00122-016-2803-2

2. Kantar MB, Sosa CC, Khoury CK, Castañeda-Álvarez NP, Achicanoy H a, Bernau V, et al. Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (Helianthus annuus L.). Front Plant Sci. 2015;6: 841. doi:10.3389/fpls.2015.00841

3. Lebeda A, Kristkova E, Kitner M, Mieslerova B, Jemelkova M, Pink DAC. Wild Lactuca species, their genetic diversity, resistance to diseases and pests, and exploitation in lettuce breeding. Eur J Plant Pathol. 2014;138: 597–640. doi:10.1007/s10658-013-0254-z

4. Hilscher J, Burstmayr H, Stoger E. Targeted modification of plant genomes for precision crop breeding. Biotechnol J. 2016;12: 1600173. doi:10.1002/biot.201600173

5. Kamthan A, Chaudhuri A, Kamthan M, Datta A. Small RNAs in plants: recent development and application for crop improvement. Front Plant Sci. 2015;6: 1–17. doi:10.3389/fpls.2015.00208

6. Jablonka E, Raz G. Transgenerational Epigenetic Inheritance: Prevalence, Mechanisms, and Implications for the Study of Heredity and Evolution. Q Rev Biol. 2009;84: 131–176. doi:10.1017/CBO9781107415324.004

7. Tal O, Kisdi E, Jablonka E. Epigenetic contribution to covariance between relatives. Genetics. 2010;184: 1037–1050. doi:10.1534/genetics.109.112466

8. Verhoeven KJF, Jansen JJ, Van Dijk PJ, Biere A. Stress-induced DNA methylation changes and their heritability in asexual dandelions. New Phytol. 2010;185: 1108–1118. doi:10.1111/j.1469-8137.2009.03121.x

9. Yi C, Zhang S, Liu X, Bui HTN, Hong Y. Does epigenetic polymorphism contribute to phenotypic variances in Jatropha curcas L.? BMC Plant Biol. 2010;10: 259. doi:10.1186/1471-2229-10-259

10. Riggs AD, Martienssen RA, Russo VE. Introduction. Epigenetic mechanisms of gene regulation. 1996. pp. 0–4. Available: https://cshmonographs.org/index.php/monographs/issue/view/087969490.32

11. Lister R, Ecker JR. Finding the fifth base: Genome-wide sequencing of cytosine methylation. Genome Res. 2009;19: 959–966. doi:10.1101/gr.083451.108

12. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome Regulation by Polycomb and Trithorax Proteins. Cell. 2007;128: 735–745. doi:10.1016/j.cell.2007.02.009

13. True HL, Berlin I, Lindquist SL. Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. Nature. England; 2004;431: 184–187. doi:10.1038/nature02885

14. Chakrabortee S, Byers JS, Jones S, Garcia DM, Bhullar B, Chang A, et al. Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. Cell. Elsevier; 2017;167: 369–381.e12. doi:10.1016/j.cell.2016.09.017

15. True HL, Lindquist SL. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. Nature. England; 2000;407: 477–483. doi:10.1038/35035005

16. Elkonin L a, Tsvetova MI. Heritable Effect of Plant Water Availability Conditions on Restoration of Male Fertility in the "9E" CMS-Inducing Cytoplasm of Sorghum. Front Plant Sci. 2012;3: 91. doi:10.3389/fpls.2012.00091

17. Bernatavichute Y V., Zhang X, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One. 2008;3. doi:10.1371/journal.pone.0003156

18. Satgé C, Moreau S, Sallet E, Lefort G, Auriac M-C, Remblière C, et al. Reprogramming of DNA methylation is critical for nodule development in Medicago truncatula. Nat Plants. Macmillan Publishers Limited; 2016;2: 16166. Available: http://dx.doi.org/10.1038/nplants.2016.166

19. Kawanabe T, Ishikura S, Miyaji N, Sasaki T, Wu LM, Itabashi E, et al. Role of DNA methylation in hybrid vigor in Arabidopsis thaliana. Proc Natl Acad Sci U S A. 2016;113: E6704–E6711. doi:10.1073/pnas.1613372113

20. Shen H, He H, Li J, Chen W, Wang X, Guo L, et al. Genome-Wide Analysis of DNA Methylation and Gene Expression Changes in Two Arabidopsis Ecotypes and Their Reciprocal Hybrids. Plant Cell. 2012;24: 875–892. doi:10.1105/tpc.111.094870

21. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11: 204–20. doi:10.1038/nrg2719

22. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462: 315–22. doi:10.1038/nature08514

23. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar a. H, et al. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell. 2008;133: 523–536. doi:10.1016/j.cell.2008.03.029

24. Heard E, Martienssen RA. Transgenerational epigenetic inheritance: Myths and mechanisms. Cell. 2014. doi:10.1016/j.cell.2014.02.045

25. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452: 215–219. doi:10.1038/nature06745

26. Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, et al. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A. 2010;107: 8689–8694. doi:10.1073/pnas.1002720107

27. Zhong S, Fei Z, Chen Y-R, Zheng Y, Huang M, Vrebalov J, et al. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat Biotechnol. 2013;31: 154–159.

28. Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, et al. CHH islands : de novo DNA methylation in near-gene chromatin regulation in maize. Genome Res. 2013;23: 628–637. doi:10.1101/gr.146985.112.as

29. Schmitz RJ, He Y, Valde O, Khan SM, Joshi T, Urich MA, et al. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res. 2013;23: 1663–1674. doi:10.1101/gr.152538.112.12

30. Seymour DK, Koenig D, Hagmann JJ, Becker C, Weigel D. Evolution of DNA Methylation Patterns in the Brassicaceae is Driven by Differences in Genome Organization. PLoS Genet. 2014;10: e1004785. doi:10.1371/journal.pgen.1004785

31. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, et al. Transgenerational epigenetic instability is a source of novel methylation variants. Science. 2011;334: 369–73. doi:10.1126/science.1212959

32. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, et al. Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature. 2011;480: 245–249. doi:10.1038/nature10555

33. Hagmann J, Becker C, Müller J, Stegle O, Meyer RC, Wang G, et al. Century-scale Methylome Stability in a Recently Diverged Arabidopsis thaliana Lineage. PLoS Genet. 2015;11: e1004920. doi:10.1371/journal.pgen.1004920

34. Zhao L, Sun M-A, Li Z, Bai X, Yu M, Wang M, et al. The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. 2014; 1296–1307. doi:10.1101/gr.163147.113.Freely

35. Yang X, Kundariya H, Xu Y-Z, Sandhu A, Yu J, Hutton SF, et al. MutS HOMOLOG1-Derived Epigenetic Breeding Potential in Tomato. Plant Physiol. 2015;168: 222–232. doi:10.1104/pp.15.00075

36. Soppe WJJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, et al. The Late Flowering Phenotype of fwa Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. Mol Cell. 2000;6: 791–802.

37. Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, et al. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science. 2004;303: 521–523. doi:10.1126/science.1089835

38. Castelletti S, Tuberosa R, Pindo M, Salvi S. A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL Vgt1. G3 (Bethesda). 2014;4: 805–12. doi:10.1534/g3.114.010686

39. Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, et al. A transposon-induced epigenetic change leads to sex determination in melon. Nature. Macmillan Publishers Limited. All rights reserved; 2009;461: 1135–1138. Available: http://dx.doi.org/10.1038/nature08498

40. Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, et al. Widespread dynamic DNA methylation in response to biotic stress. Proc Natl Acad Sci. 2012;109: E2183–E2191. doi:10.1073/pnas.1209329109

41. Yu A, Lepère G, Jay F, Wang J, Bapaume L, Wang Y, et al. Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. Proc Natl Acad Sci U S A. 2013;110: 2389–94. doi:10.1073/pnas.1211757110

42. Rasmann S, De Vos M, Casteel CL, Tian D, Halitschke R, Sun JY, et al. Herbivory in the Previous Generation Primes Plants for Enhanced Insect Resistance. Plant Physiol. 2012;158: 854–863. doi:10.1104/pp.111.187831

43. Eichten SR, Springer NM. Minimal evidence for consistent changes in maize DNA methylation patterns following environmental stress. Front Plant Sci. 2015;6: 308. doi:10.3389/fpls.2015.00308

44. Karan R, DeLeon T, Biradar H, Subudhi PK. Salt stress induced variation in DNA methylation pattern and its influence on gene expression in contrasting rice genotypes. PLoS One. 2012;7. doi:10.1371/journal.pone.0040203

45. Ferreira LJ, Azevedo V, Maroco JJ, Oliveira MM, Santos AP. Salt Tolerant and Sensitive Rice Varieties Display Differential Methylome Flexibility under Salt Stress. PLoS One. 2015;10: e0124060. doi:10.1371/journal.pone.0124060

46. Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y, et al. Transgenerational adaptation of Arabidopsis to stress requires DNA methylation and the function of dicer-like proteins. PLoS One. 2010;5. doi:10.1371/journal.pone.0009514

47. Wada Y, Miyamoto K, Kusano T, Sano H. Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in tobacco plants. Mol Genet Genomics. 2004;271: 658–666. doi:10.1007/s00438-004-1018-4

48. Wang W-S, Pan Y-J, Zhao X-Q, Dwivedi D, Zhu L-H, Ali J, et al. Drought-induced site-specific DNA methylation and its association with drought tolerance in rice (Oryza sativa L.). J Exp Bot. 2011;62: 1951–60. doi:10.1093/jxb/erq391

49. Medvedeva YA, Khamis AM, Kulakovskiy I V, Ba-Alawi W, Bhuyan MSI, Kawaji H, et al. Effects of cytosine methylation on transcription factor binding sites. BMC Genomics. 2014;15: 119. doi:10.1186/1471-2164-15-119

50. Guo H, Hu B, Yan L, Yong J, Wu Y, Gao Y, et al. DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. Cell Res. Nature Publishing Group; 2017;27: 165–183. doi:10.1038/cr.2016.128

51. Liu R, How-Kit A, Stammitti L, Teyssier E, Rolin D, Mortain-Bertrand A, et al. A DEMETER-like DNA demethylase governs tomato fruit ripening. Proc Natl Acad Sci U S A. 2015;112: 10804–9. doi:10.1073/pnas.1503362112

52. Williams BP, Pignatta D, Henikoff S, Gehring M. Methylation-Sensitive Expression of a DNA Demethylase Gene Serves As an Epigenetic Rheostat. PLOS Genet. 2015;11: e1005142. doi:10.1371/journal.pgen.1005142

53. Wang J, Marowsky NC, Fan C. Divergence of gene body DNA methylation and evolution of plant duplicate genes. PLoS One. 2014;9. doi:10.1371/journal.pone.0110357

54. Li X, Zhu J, Hu F, Ge S, Ye M, Xiang H, et al. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. BMC Genomics. BMC Genomics; 2012;13: 300. doi:10.1186/1471-2164-13-300

55. Li Q, Song J, West PT, Zynda G, Eichten SR, Vaughn MW, et al. Examining the causes and consequences of context-specific differential DNA methylation in maize. Plant Physiol. 2015; pp.00052.2015. doi:10.1104/pp.15.00052

56. Liu S, Yeh CT, Ji T, Ying K, Wu H, Tang HM, et al. Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. PLoS Genet. 2009;5. doi:10.1371/journal.pgen.1000733

57. Flores K, Wolschin F, Corneveaux JJ, Allen AN, Huentelman MJ, Amdam G V. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genomics. BMC Genomics; 2012;13: 480. doi:10.1186/1471-2164-13-480

58. Regulski M, Lu Z, Kendall J, Donoghue MTA, Reinders J, Llaca V, et al. The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Res. 2013;23: 1651–1662. doi:10.1101/gr.153510.112

59. Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, et al. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A. 2010;107: 8689–94. doi:10.1073/pnas.1002720107

60. Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science. 2012;337: 1360–4. doi:10.1126/science.1224839

61. Aina R, Sgorbati S, Santagostino A, Labra M, Ghiani A, Citterio S. Specific hypomethylation of DNA is induced by heavy metals in white clover and industrial hemp. Physiol Plant. Munksgaard International Publishers; 2004;121: 472–480. doi:10.1111/j.1399-3054.2004.00343.x

62. Boyko A, Kathiria P, Zemp FJ, Yao Y, Pogribny I, Kovalchuk I. Transgenerational changes in the genome stability and methylation in pathogen-infected plants: (Virus-induced plant genome instability). Nucleic Acids Res. 2007;35: 1714–1725. doi:10.1093/nar/gkm029

63. Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I. Stress-induced activation of heterochromatic transcription. PLoS Genet. 2010;6: 1–11. doi:10.1371/journal.pgen.1001175

64. Hashida S-N, Uchiyama T, Martin C, Kishima Y, Sano Y, Mikami T. The temperature-dependent change in methylation of the Antirrhinum transposon Tam3 is controlled by the activity of its transposase. Plant Cell. 2006;18: 104–118. doi:10.1105/tpc.105.037655

65. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res. 2009;19: 1419–1428. doi:10.1101/gr.091678.109

66. Ahmed I, Sarazin A, Bowler C, Colot V, Quesneville H. Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Res. 2011;39: 6919–6931. doi:10.1093/nar/gkr324

67. Eichten SR, Ellis NA, Makarevitch I, Yeh CT, Gent JI, Guo L, et al. Spreading of Heterochromatin Is Limited to Specific Families of Maize Retrotransposons. PLoS Genet. 2012;8. doi:10.1371/journal.pgen.1003127

68. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, et al. Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress. PLoS Genet. 2015;11. doi:10.1371/journal.pgen.1004915

69. Le T-N, Schumann U, Smith NA, Tiwari S, Au P, Zhu Q-H, et al. DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis. Genome Biol. 2014;15: 458. doi:10.1186/s13059-014-0458-3

70. Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, et al. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. Elife. 2015;4: e09343. doi:10.7554/eLife.09343.001

71. Carone BR, Fauquier L, Habib N, Shea JM, Hart CE, Li R, et al. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. Cell. Elsevier Inc.; 2010;143: 1084–1096. doi:10.1016/j.cell.2010.12.008

72. Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, et al. DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat Commun. 2014;5: 5592. doi:10.1038/ncomms6592

73. Ng S-F, Lin RCY, Laybutt DR, Barres R, Owens JA, Morris MJ. Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. Nature. England; 2010;467: 963–966. doi:10.1038/nature09491

74. Feinberg AP, Irizarry RA. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci U S A. 2010;107 Suppl: 1757–1764. doi:10.1073/pnas.0906183107

75. Latzel V, Allan E, Bortolini Silveira A, Colot V, Fischer M, Bossdorf O. Epigenetic diversity increases the productivity and stability of plant populations. Nat Commun. 2013;4: 2875. doi:10.1038/ncomms3875

76. Economic Research Service. U.S. Department of Agriculture (USDA). In: U.S. Lettuce Statistics 2011. 2011.

77. Christopoulou M, McHale LK, Kozik A, Reyes-Chin Wo S, Wroblewski T, Michelmore RW. Dissection of Two Complex Clusters of Resistance Genes in Lettuce (Lactuca sativa). Mol Plant Microbe Interact. United States; 2015;28: 751–765. doi:10.1094/MPMI-06-14-0175-R

78. Lindqvist K. On the Origin of Cultivated Lettuce. Hereditas. 1960;49: 319–350. doi:10.2307/2441230

79. De Vries IM. Origin and domestication of Lactuca sativa L. Genet Resour Crop Evol. 1997;44: 165–164. doi:10.1023/A:1008611200727

80. Kesseli R V., Ochoa O, Michelmore RW. Variation at RFLP loci in Lactuca spp. and the origin of cultivated lettuce (L. sativa). Genome. 1991;34: 430–436.

81. Hill M, Witsenboer M, Zabeau PV, Kesseli R V., Michelmore RW. Amplified fragment length polymorphisms (AFLPs) as a tool for assessing the genetic relationships in Lactuca spp. Theor Appl Genet. 1996;93: 1202–1210.

82. Sturtevant E. A study of garden lettuce. Am Nat. 1886;20: 230–233.

83. National Plant Data Team. The PLANTS Database [Internet]. Greensboro, NC: Natural Resource Conservation Service, United States Department of Agriculture; 2016. Available: http://plants.usda.gov

84. Akimoto K, Katakami H, Kim H-J, Ogawa E, Sano CM, Wada Y, et al. Epigenetic inheritance in rice plants. Ann Bot. 2007;100: 205–17. doi:10.1093/aob/mcm110

85. Buscaill P, Rivas S. Transcriptional control of plant defence responses. Curr Opin Plant Biol. Elsevier Ltd; 2014;20: 35–46. doi:10.1016/j.pbi.2014.04.004

86. Sekhwal M, Li P, Lam I, Wang X, Cloutier S, You F. Disease Resistance Gene Analogs (RGAs) in Plants. Int J Mol Sci. 2015;16: 19248–19290. doi:10.3390/ijms160819248

87. Dangl JL, McDowell JM. Two modes of pathogen recognition by plants. Proc Natl Acad Sci U S A. 2006;103: 8575–8576. doi:10.1073/pnas.0603183103

88. Farrara BF, Ilott TW, Michelmore RW. Genetic analysis of factors for resistance to downy mildew (Bremia lactucae) in species of lettuce (Lactuca sativa and L. seriola). Plant Pathol. 1987;36: 499–514.

89. Meyers BC, Shen K a, Rohani P, Gaut BS, Michelmore RW. Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell. 1998;10: 1833–1846. doi:10.1105/tpc.10.11.1833

90. Uwimana B, Smulders MJM, Hooftman D a. P, Hartman Y, van Tienderen PH, Jansen J, et al. Hybridization between crops and wild relatives: the contribution of cultivated lettuce to the vigour of crop–wild hybrids under drought, salinity and nutrient deficiency conditions. Theor Appl Genet. 2012;125: 1097–1111. doi:10.1007/s00122-012-1897-4

91. Shim CK, Kim MJ, Kim YK, Jee HJ. Evaluation of lettuce germplasm resistance to gray mold disease for organic cultivations. Plant Pathol J. 2014;30: 90–95. doi:10.5423/PPJ.NT.07.2013.0064

92. Kerbiriou PJ, Maliepaard CA, Stomph TJ, Koper M, Froissart D, Roobeek I, et al. Genetic Control of Water and Nitrate Capture and Their Use Efficiency in Lettuce (Lactuca sativa L.). Front Plant Sci. 2016;7: 1–14. doi:10.3389/fpls.2016.00343

93. Mikel MA. Genealogy of contemporary North American lettuce. HortScience. 2007;42: 489–493.

94. Wilkins O, Bräutigam K, Campbell MM. Time of day shapes Arabidopsis drought transcriptomes. Plant J. 2010;63: 715–727. doi:10.1111/j.1365-313X.2010.04274.x

95. Wilkins O, Waldron L, Nahal H, Provart NJ, Campbell MM. Genotype and time of day shape the Populus drought response. Plant J. 2009;60: 703–715. doi:10.1111/j.1365-313X.2009.03993.x

96. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014; doi:10.1093/bioinformatics/btu170

97. Renaud G, Stenzel U, Kelso J. leeHom : adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Res. 2014;42: e141. doi:10.1093/nar/gku699

98. Krueger F, Andrews SR. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27: 1571–1572. doi:10.1093/bioinformatics/btr167

99. Picard Toolkit [Internet]. Available: http://broadinstitute.github.io/picard/

100. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Meth. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2012;9: 357–359. Available: http://dx.doi.org/10.1038/nmeth.1923

101. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; doi:10.1093/bioinformatics/btq033

102. Smit A, Hubley R, Green P. Repeat Masker Open-4.0 [Internet]. Available: http://www.repeatmasker.org

103. Park Y, Figueroa ME, Rozek LS, Sartor M a. MethylSig: a whole genome DNA methylation analysis pipeline. Bioinformatics. 2014;30: 1–8. doi:10.1093/bioinformatics/btu339

104. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nat Commun. Nature Publishing Group; 2016;7: 10478. doi:10.1038/ncomms10478

105. Teschendorff AE, Widschwendter M. Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. Bioinformatics. 2012;28: 1487–1494. doi:10.1093/bioinformatics/bts170

106. Wang H, Beyene G, Zhai J, Feng S, Fahlgren N, Taylor NJ, et al. CG gene body DNA methylation changes and evolution of duplicated genes in cassava. Proc Natl Acad Sci U S A. 2015;112: 13729–13734. doi:10.1073/pnas.1519067112

107. Song Q-X, Lu X, Li Q, Chen H, Hu X-Y, Ma B, et al. Genome-Wide Analysis of DNA Methylation in Soybean. Mol Plant. © 2013 The Authors. All rights reserved.; 2013;6: 1961–1974. doi:10.1093/mp/sst123

108. Christopoulou M, Wo SR-C, Kozik A, McHale LK, Truco M-J, Wroblewski T, et al. Genome-Wide Architecture of Disease Resistance Genes in Lettuce. G3 (Bethesda). 2015;5: 2655–69. doi:10.1534/g3.115.020818

109. Teschendorff AE, Jones A, Widschwendter M. Stochastic epigenetic outliers can define field defects in cancer. BMC Bioinformatics. 2016; 178. doi:10.1186/s12859-016-1056-z

110. Alonso C, Perez R, Bazaga P, Herrera CM. Global DNA cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms. Front Genet. 2015;6: 1–9. doi:10.3389/fgene.2015.00004

111. Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim K Do, Li Q, et al. Widespread natural variation of DNA methylation within angiosperms. doi:10.1186/s13059-016-1059-0

112. Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet. England; 2014;15: 394–408. doi:10.1038/nrg3683

113. Zhong X, Wang Y, Liu X, Gong L, Ma Y, Qi B, et al. DNA methylation polymorphism in annual wild soybean ( Glycine soja Sieb. et Zucc.) and cultivated soybean ( G. max L. Merr.). Can J Plant Sci. 2009;89: 851–863. doi:10.4141/CJPS08215

114. Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, et al. Epigenetic and genetic influences on DNA methylation variation in maize populations. [Internet]. The Plant cell. 2013. doi:10.1105/tpc.113.114793

115. Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu S, et al. Heritable Epigenetic Variation among Maize Inbreds. PLoS Genet. 2011;7: e1002372. doi:10.1371/journal.pgen.1002372

116. Schmitz RJ, He Y, Valdés-lópez O, Res G, Gent JI, Ellis N a, et al. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. 2013; 1663–1674. doi:10.1101/gr.152538.112

117. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, et al. Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature. 2011;480: 245–249. doi:10.1038/nature10555

118. Rambani A, Rice JH, Liu J, Lane T, Ranjan P, Mazarei M, et al. The Methylome of Soybean Roots during the Compatible Interaction with the Soybean Cyst Nematode. Plant Physiol. 2015;168: 1364–1377. doi:10.1104/pp.15.00826

119. Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet. Nature Publishing Group; 2013;14: 840–852. doi:10.1038/nrg3605

120. Smith JWM. Recurring off-types in lettuce: Their significance in plant breeding and seed production. Theor Appl Genet. 1977;50: 79–87.

121. Stroud H, Ding B, Simon SA, Feng S, Bellizzi M, Pellegrini M, et al. Plants regenerated from tissue culture contain stable epigenome changes in rice. Elife. 2013;2013: 1–14. doi:10.7554/eLife.00354

122. Vining K, Pomraning KR, Wilhelm LJ, Ma C, Pellegrini M, Di Y, et al. Methylome reorganization during in vitro dedifferentiation and regeneration of Populus trichocarpa. BMC Plant Biol. 2013;13: 92. doi:10.1186/1471-2229-13-92

123. Peschke VM, Phillips RL, Gengenbach BG. Discovery of transposable element activity among progeny of tissue culture--derived maize plants. Science. United States; 1987;238: 804–807. doi:10.1126/science.238.4828.804

124. Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL. DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A. 2007;104: 6752–6757. doi:10.1073/pnas.0701861104

125. Gehring M, Bubb KL, Henikoff S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science. 2009;324: 1447–1451. doi:10.1126/science.1171609

126. Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu J-K, et al. Regulatory link between DNA methylation and active demethylation in Arabidopsis. Proc Natl Acad Sci U S A. 2015;112: 3553–7. doi:10.1073/pnas.1502279112

127. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. Macmillian Magazines Ltd.; 2000;408: 796–815. Available: http://dx.doi.org/10.1038/35048692

128. Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science. United States; 2002;296: 92–100. doi:10.1126/science.1068275

129. Jacquemin J, Bhatia D, Singh K, Wing R a. The International Oryza Map Alignment Project: Development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. Curr Opin Plant Biol. Elsevier Ltd; 2013;16: 147–156. doi:10.1016/j.pbi.2013.02.014

130. Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science. United States; 2002;296: 79–92. doi:10.1126/science.1068037

131. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, et al. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. file///Users/tgulick/Downloads/citations (5).nbibNature Commun. England; 2014;5: 3930. doi:10.1038/ncomms4930

132. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. England; 2012;485: 635–641. doi:10.1038/nature11119

133. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature. England; 2010;463: 178–183. doi:10.1038/nature08670

134. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. United States; 2009;326: 1112–1115. doi:10.1126/science.1178534

135. Gugger PF, Fitz-Gibbon S, Pellegrini M, Sork VL. Species-wide patterns of DNA methylation variation in Quercus lobata and its association with climate gradients. Mol Ecol. 2016;

136. Gulick TA, Morey SH, Avery A, Kesseli R V. Conservation and variation in DNA methylation in Lactuca sativa and Lactuca serriola. Prep. 2016;

137. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170

138. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 1995;57: 289–300. doi:10.2307/2346101

139. Gulick T, Avery A, Morey S, Kesseli R. Reduced representation bisulfite sequencing of Lactuca serriola and Lactuca sativa Salinas with different family histories of nutrient deprivation. Manuscr Prep. 2016;

140. Jablonka E, Lamb MJ. The inheritance of acquired epigenetic variations. J Theor Biol. ENGLAND; 1989;139: 69–83.

141. Agrawal AA, Strauss SY, Stout MJ. Costs of Induced Responses and Tolerance to Herbivory in Male and Female Fitness Components of Wild Radish. Evolution (N Y). 1999;53: 1093–1104. doi:10.2307/2640814

142. Molinier J, Ries G, Zipfel C, Hohn B. Transgeneration memory of stress in plants. Nature. 2006;442: 1046–1049. Available: http://dx.doi.org/10.1038/nature05022

143. Holeski LM. Within and between generation phenotypic plasticity in trichome density of Mimulus guttatus. J Evol Biol. Switzerland; 2007;20: 2092–2100. doi:10.1111/j.1420-9101.2007.01434.x

144. Gao L, Geng Y, Li B, Chen J, Yang J. Genome-wide DNA methylation alterations of Alternanthera philoxeroides in natural and manipulated habitats: Implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. Plant, Cell Environ. 2010;33: 1820–1827. doi:10.1111/j.1365-3040.2010.02186.x

145. Hess M, Barralis G, Bleiholder H, Buhr L, Eggers T, Hack H, et al. Use of the extended BBCH-scale general for the descriptions of the growth stages of mono- and dicotyledonous weed species. Weed Res. 1997;37: 433–441.

146. Gulick TA, Morey SH, Avery A, Kesseli R V. Reduced representation bisulfite sequencing of a highly repetitive plant genome. Prep. 2016;

147. AE T, A J, M W. Stochastic epigenetic outliers can define field defects in cancer. BMC Bioinformatics. BMC Bioinformatics; 2016; 178. doi:10.1186/s12859-016-1056-z

148. Paape T, Zhou P, Branca A, Briskine R, Young N, Tiffin P. Fine-scale population recombination rates, hotspots, and correlates of recombination in the Medicago truncatula genome. Genome Biol Evol. 2012; doi:10.1093/gbe/evs046

149. Arbeithuber B, Betancourt AJ, Ebner T, Tiemann-Boege I, Hurst LD. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. doi:10.1073/pnas.1416622112

150. Yelina NE, Lambing C, Hardcastle TJ, Zhao X, Santos B, Henderson IR. DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. Genes Dev. 2015;29: 2183–2202. doi:10.1101/gad.270876.115

151. Shilo S, Melamed-Bessudo C, Dorone Y, Barkai N, Levy AA. DNA Crossover Motifs Associated with Epigenetic Modifications Delineate Open Chromatin Regions in Arabidopsis. doi:10.1105/tpc.15.00391

152. Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, et al. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. Elife. 2015;4: e09343. doi:10.7554/eLife.09343

153. Kou HP, Li Y, Song XX, Ou XF, Xing SC, Ma J, et al. Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (Oryza sativa L.). J Plant Physiol. 2011;168: 1685–1693. doi:10.1016/j.jplph.2011.03.017

154. Ou X, Zhang Y, Xu C, Lin X, Zang Q, Zhuang T, et al. Transgenerational Inheritance of Modified DNA Methylation Patterns and Enhanced Tolerance Induced by Heavy Metal Stress in Rice (Oryza sativa L.). PLoS One. 2012;7. doi:10.1371/journal.pone.0041143

155. Crisp PA, Ganguly D, Eichten SR, Borevitz JO, Pogson BJ. Reconsidering plant memory: Intersections between stress recovery, RNA turnover, and epigenetics. 2016; doi:10.1126/sciadv.1501340

156. Thrall PH, Bever JD, Burdon JJ. Evolutionary change in agriculture: The past, present and future. Evol Appl. 2010; doi:10.1111/j.1752-4571.2010.00155.x

157. Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, et al. Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell. 2013;25: 2783–2797. doi:10.1105/tpc.113.114793

158. Brencic A, Winans SC, Colonization P. Detection of and Response to Signals Involved in Host-Microbe Interactions by Plant-Associated Bacteria. 2005;69: 155–194. doi:10.1128/MMBR.69.1.155

159. Parke D, Rivelli M, Ornston LN. Chemotaxis to aromatic and hydroaromatic acids: Comparison of Bradyrhizobium japonicum and Rhizobium trifolii. J Bacteriol. 1985;163: 417–422.

160. Mandal SM, Chakraborty D, Dey S. Phenolic acids act as signaling molecules in plant-microbe symbioses. Plant Signal Behav. 2010;5: 359–68. doi:10.4161/psb.5.4.10871

161. Tsuchiya T, Eulgem T. An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. Proc Natl Acad Sci U S A. 2013;110: E3535-43. doi:10.1073/pnas.1312545110

162. Da K, Nowak J, Flinn B. Potato cytosine methylation and gene expression changes induced by a beneficial bacterial endophyte, Burkholderia phytofirmans strain PsJN. Plant Physiol Biochem. Elsevier Masson SAS; 2012;50: 24–34. doi:10.1016/j.plaphy.2011.09.013

163. Berg G. Plant--microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. Appl Microbiol Biotechnol. 2009;84: 11–18. doi:10.1007/s00253-009-2092-7